Interdisciplinary experiments on reciprocal fairness, altruism and the sunk-cost bias

Document status and date:
Published: 01/01/2021

DOI:
10.26481/dis.20210521mn

Document Version:
Publisher's PDF, also known as Version of record

Interdisciplinary experiments on reciprocal fairness, altruism, and the sunk-cost bias.

The author acknowledges financial support from Maastricht University, the Graduate School of Business and Economics at Maastricht University and from the Maastricht University Center for Neuroeconomics (MU-CEN).

This book was typeset by the author using LaTeX.The cover art is the work of Andrea Miotti.

Interdisciplinary experiments on reciprocal fairness, altruism, and the sunk-cost bias.

Dissertation

to obtain the degree of Doctor at Maastricht University
on the authority of the Rector Magnificus Prof.dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,

to be defended in public on

Friday, 21 May 2021 at 10:00 hours.

by

Marcello Negrini

Supervisor: Prof. Dr. Arno Riedl
Co-Supervisor: Dr. Matthias Wibral

Assessment Committee:

Prof. dr. Elisabeth Brüggen (Chair)
Prof. dr. Carolyn Declerck, University of Antwerp
Dr. Claudia Civai, London South Bank University
Dr. Peter Werner

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *"Perhaps I can best illustrate the meaning of my thoughts by going back to Oppenheimer's felicitous metaphor of the house called "science". I would like to see us build a new room in that vast and rambling structure. This room, like the others, would have no door and over the entrance would be the words, thought, reflection, contemplation. It would have no tables with instruments, no whirring machinery. There would be no sound except the soft murmur of words carrying the thoughts of men [and women] in the room. It would be a Commons Room to which men [and women] would drift from those rooms marked geology, anthropology, taxonomy, technology, biology, paleontology, logic, mathematics, psychology, linguistic, and many others. Indeed, from without the walls of the House would come poets and artists. All these would drop in and linger. This room would have great windows, the vistas our studies have opened. Men [and women] singly or together would from time to time walk to those windows to gaze out on the landscape beyond. This landscape in all Its beauty, sometimes gentle, sometimes terrible, cannot be seen fully by any one of the occupants of the room. Indeed, It cannot be known fully by a whole generation of men [and women]. Explorers of each generation travel into its unknown recesses and, with luck, return to share their discoveries with us. So the life of the new room would go on - thought, reflection, contemplation - as the explorers bring back their discoveries to share with the room's occupants. This landscape that we gaze on and try to understand is an epic portion of the human experience."*[1]

Real-world problems and challenges are rarely confined by the artificial boundaries of any single academic discipline and approach but require diverse thoughts, reflection, and contemplation (You, 2017). Moving beyond the artificial fragmentation of knowledge and joining a common room in the *house of science* can be a fruitful endeavor for the pursuit of truth. Indeed, collaboration among disciplinary approaches is currently one of the most promoted goals at the level of scientific administration, both in the natural and social sciences (MacLeod, 2018). It is often argued that collaboration among multiple disciplines is needed to meet the demands of many societal, environmental, industrial, and scientific problems that cannot be adequately

[1] The citation is from Margaret Mead's Book "Blackbery Winter, as reported in Stember (1991).

addressed by a single discipline alone (Bridle et al., 2013). Several policy reports put a strong emphasis on interdisciplinary research and promote the institutional changes needed to support it. For instance, the Royal Society argues that: “Many of the major challenges that society faces today will require solutions developed through interdisciplinary research and cross-disciplinary collaboration. Improving support for and addressing the barriers to this work could contribute to major scientific breakthroughs at the interface of disciplines, develop new technologies, and ultimately support the economy and develop novel solutions to societal challenges”.[2] Science Europe, the association of European funding organizations, also claims that the “key to future scientific breakthroughs lies in interdisciplinary research”. Finally, the Global Research Council, which is constituted by funders such as the U.S. National Science Foundation, the Research Councils UK, and the Chinese Academy of Sciences, also produced its own report on interdisciplinary research.

For most of modern history, science and knowledge followed a pyramidal structure, where subordinate elements moved under the direction of a leading science such as theology, law, arts, and medicine (Salter and Hearn, 1997). It has been only recently, mostly during the nineteenth and twentieth centuries, that our modern individual disciplines emerged and claimed their independence. The rapidly changing and urbanizing society generated new problems and challenges. This period saw the formation of new categories of knowledge. Social sciences, including anthropology, economics, and political science emerged. Soon after that, the subdisciplines of sociology, psychology, and history followed (Easton, 1991).

This conceptual specialization has yielded remarkable gains in knowledge, but fragmentation resulted in some disciplines to become so isolated and able to understand and contribute to only a small part of a complex problem. The dissatisfaction with the compartmentalization of disciplines and knowledge has its origin in the idea that any field is enriched by theories, concepts, and methods from other fields (Ross and Staw, 1993). As a result, during the last decades, there have been clear signs indicating that clusters of disciplines are moving again toward unity, but not following a medieval system of subordination of certain disciplines to others. Instead, the goal is the creation of a coexisting open-minded coalition motivated by the idea that knowledge can be understood and advanced through interdisciplinary work (Stember, 1991).

In such a collaborative model, the integration creates a community of learning that fosters creative intellectual inquiry and facilitates solving related problems at a different level of analysis and from different perspectives. One consequence of this trend is the regrouping of disciplines according to the field of study. For instance, economics involves mathematics, sociology, business administration, etc. Another tendency has been the creation of new disciplines that merge traditional boundaries. Behavioral and neuro economics are recent examples of this effort, as they are disciplines that propose merging the economic approach with insights from psychology, biology, and neuroscience with the goal to better study human decision making.

[2]The Royal Society (2015): Response to the British Academy’s call for evidence on ‘Interdisciplinarity’. The report is available at https://royalsociety.org

1. Different disciplinarities

In recent years, it became evident that there is a growing emphasis on the development of teamwork that requires a multidisciplinary approach (Choi and Pak, 2007). There is also more attention from funding agencies which often call for research that involves multiple disciplines (Choi and Pak, 2007) and pressure from universities to establish multiple disciplinary departments and teaching programs (Slatin et al., 1991). However, there is still little consensus on the definition of the terms used to define an effort that goes beyond each discipline's boundaries. Terms like multidisciplinary, interdisciplinary, and transdisciplinary are often used interchangeably and are ambiguously defined (Whitfield and Reid, 2004). In what follows, I briefly describe to the unfamiliar reader the main differences between these approaches and I will later contextualize the research presented in this thesis using this framework.

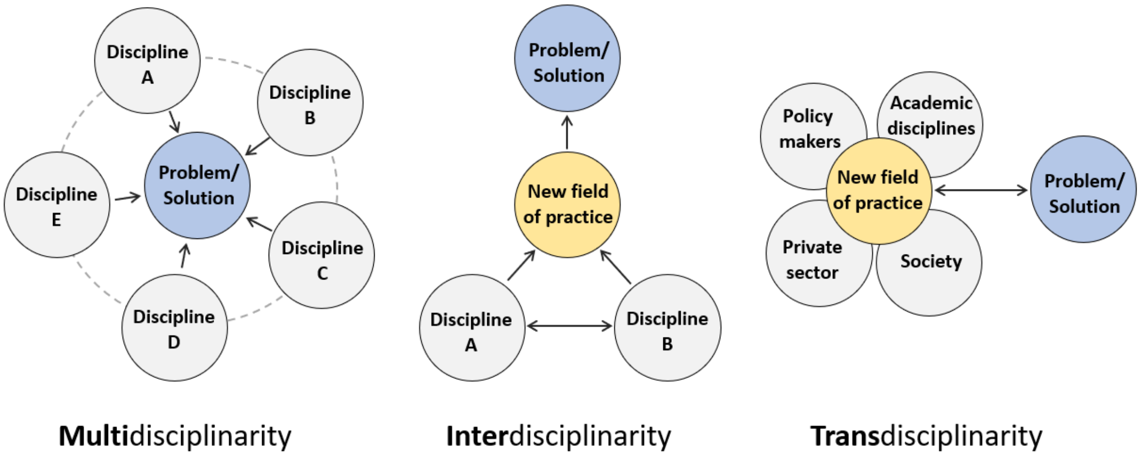


Figure 1.1: **Different levels of disciplinarity**

A comparison of multidisciplinary, interdisciplinary, and transdisciplinary approaches. The image is adapted from McPhee et al. (2018)

- **Intradisciplinarity**. Indicating work that remains within the boundaries of a single discipline and that addresses a specific problem of that discipline.

- **Multidisciplinary.** This approach is a level higher compared to intradisciplinarity, as it draws from different disciplines. Following the definition proposed by Klein (1990), multidisciplinarity is a process that juxtaposes disciplines in an additive fashion, but that avoids integration. Thus, the individual disciplinary perspective is not changed, but only contrasted. Thus, the independence of each discipline is preserved and the collaboration with different disciplines does not result in changes in the existing structures. An example of such an approach is scholars from different disciplines working together, but each referring to the knowledge of their discipline. For instance, an ancient song can be studied not only from the perspective of musicology but also within anthropology and linguistics. Multidisciplinary teams work in parallel or sequentially from their specific disciplinary base to address a common problem.

- **Interdisciplinarity.** While multidisciplinarity requires little interaction or collaboration across disciplines, interdisciplinarity acknowledges the limitation of the compartmentalization of each individual discipline and integrates knowledge and methods from different disciplines, using a real synthesis of approaches. This results in an effort that synthesizes and harmonizes links between disciplines into a coordinated and coherent whole (Choi and Pak, 2007). The disciplines involved in an interdisciplinary effort generate a common uniform conceptual and methodological framework which transcends the structure of a single discipline and allows integration and exchange between different schools of thought which synthesizes two or more disciplines and establish a new level of discourse and integration of knowledge (Klein, 1990). The discipline of neuroeconomics is an example of such an approach, where the economic study of decision making incorporates insights, theories, and methods from psychology and neuroscience.

- **Transdisciplinarity.** A transdisciplinary approach is concerned with the unity of intellectual frameworks beyond the disciplinary perspectives (Reilly, 2001). Transdisciplinarity provides holistic schemes that subordinate disciplines, looking at the dynamics of whole systems Klein (1990). Etymologically, “Trans-” can be defined as “across; over; beyond; on the further side. This approach requires to consider the existing elements constituting a discipline, breaking them down into essential elements, and recombining them to form a new coding or new knowledge. Relatively to multidisciplinary and interdisciplinary approaches, a transdisciplinary approach is more than a coordinated input from multiple disciplines that work toward a shared goal and that interact with each other. The interdisciplinary approach concerns the transfer of methods from one discipline to another, allowing research to spill over disciplinary boundaries, but staying within the framework of disciplinary research. Instead, in a transdisciplinary approach, researchers from different disciplines work jointly, intending to create new conceptual, theoretical, methodological, and translational innovations that integrate and move beyond discipline-specific approaches. Solving the problem results in the generation of new insights formed via the integration of those domains that contribute back to them in an overarching unity of knowledge. Another distinctive feature of transdisciplinary research is the inclusion of various stakeholders. These are both from within academia, as well as non-academic stakeholders such as companies, non-governmental organizations (NGOs), media, and local citizens are also participating in solving the problem together (Rice, 2013).

2. Barriers and facilitators to interdisciplinary research

While many papers and policy reports advocate the importance of moving beyond an intradisciplinary approach (MacLeod, 2018), several obstacles hinder the progression towards collaboration among disciplines. For the sake of brevity, I only mention the ones that in my view are the most relevant:

- **Academic structure.** University is traditionally organized in faculties and departments. Each department focuses on providing organizational support to advance learning and research primarily in its field. Collaboration between departments or faculties could be difficult when these bodies have structures that do not align with each other. In addition, there is also a physical limit of disciplinary scholars usually located in distinct buildings on campuses.
- **Attitude and communication.** Idiosyncrasies of each individual discipline might represent an obstacle to successful collaboration as different language use: jargon and technical language do not promote an environment and terminology that is easy for everyone involved to understand. Additionally, other disciplines might be considered as less important or rigorous in their methods relative to the perceived superiority of the "home discipline".
- **Promotions and career prospects.** Within universities, funds are often allocated to departments that also grant promotions and tenure which rely on departmental contribution and publications. Research that involves multiple disciplines and departments may take longer than traditional disciplinary work. This could result in less publication in peer-reviewed journals, which is one of the main criteria for tenure in most universities. In addition, there are fewer avenues for publishing interdisciplinary research than single-discipline research.

Interdisciplinary research can thus pose particular challenges. However, Morse et al. (2007) presented several recommendations regarding interdisciplinary collaboration and research that can facilitate these endeavors. In addition to the necessary structural support from the University (e.g., vision in hiring, promotion requirements, fundings, etc.), I summarize three recommendations that I find the most important:

- **Define focal themes and research questions jointly and clearly.** A shared team vision can promote team engagement and a clear research agenda from the beginning. Common vision facilitates the integration of multiple disciplinary backgrounds.
- **Target interdisciplinary training**. To tackle cross-disciplinary differences, seminars that cover interdisciplinary topics and methods provide useful insights to team members and can help foster creativity and generate a commitment to interdisciplinary work.
- **Select team members thoughtfully and strategically.** The selection of team members is crucial for an interdisciplinary team. Members are required to be flexible and work well in dynamic settings where the boundaries of the home disciplines are redefined.

3. Overview of this thesis

Broadly speaking, the research presented in this thesis is the result of an interdisciplinary effort in the study of economic relevant decisions. This approach originates from an effort that integrates insights from economics, but that also welcomes theories and methodologies from psychology and cognitive neuroscience. The use of laboratory experiments is the common denominator underlying the studies discussed in the following chapters. Relatively to other empirical methods available to study human decision-making, one of the biggest advantages of conducting experiments – and in particular lab experiments – is the opportunity to create the environment which is best suited to investigate specific processes without losing control of unobserved characteristics. This is crucial when the aim is to establish causality between a variable and behavior, to elicit economic preferences, or test the predictions of theoretical models. There are several features of experimental research discussed in this thesis that are worth to be mentioned. All of them are crucial to allow replicability, which is an essential feature of scientific research, and to maintain control over the experiment and subjects' behavior.

- **No deception rule.** This rule ensures that participants believe the instructions. The concern here is that if participants in a study don't believe the researcher, their decisions will be unreliable.
- **Real financial incentive.** Providing incentives for choices and outcomes leads to more meaningful and reliable choice behavior. Specifically, incentives that encourage subjects to make honest and non-arbitrary choices will accurately reveal the characteristics of their preferences.
- **Randomization and anonymity.** The concerns here involve ensuring that participants in an experiment are randomly allocated to treatments. Under the random assignment, the treatment groups should not differ significantly, and it is thus possible to establish a causal relationship between the intervention and the behavioral outcome. Anonymity in experiments is used to eliminate observer and experiment demand effects and to promote behavior that would be present in private settings.

Description of the Chapters

In **Chapter 2**, *Love and retaliation*, I describe a study that investigates the roots of human justice in the context of reciprocity that takes place in social punishment and reward. This project adopts the neuroeconomics interdisciplinary approach and integrates insights from Economics with Psychology and Neuroscience. Specifically, we test the involvement of two areas of the prefrontal cortex of the human brain (the dorsolateral prefrontal cortex and the medial prefrontal cortex) in the decision to reciprocate selfish behavior with punishment and kind actions with rewards. We consider different social contexts, which vary in personal involvement (i.e., second vs. third-party) and in the cost to punish and reward (i.e., cost vs. no cost). To test the involvement of the prefrontal cortex in these situations, we use transcranial magnetic stimulation (TMS). With this technique, it is possible to stimulate brain areas from outside the head as it delivers a magnetic pulse and allows to stimulate neurons in the underlying brain area.

The study provides evidence of the involvement of these prefrontal regions in punishment and reward. The experimental design exhibits several desirable features as we are able to (1) analyze how punishment and reward relate to each other at the behavioral level, (2) discuss whether the different mechanisms associated with these prefrontal regions extend beyond the contexts in which they have been originally tested, and (3) provide evidence of the differences and similarities between punishment and reward regarding the engagement of the right dorsolateral and the medial prefrontal cortex. In addition, we provide insights into how punishment and reward are related to motives such as outcome maximization, norm enforcement, and reciprocity.

In **Chapter 3**, *Intertemporal Social Preferences*, I continue my investigation on social preferences, and I discuss an experimental study on altruism with time delays. Despite the importance of these decisions, intertemporal and social domains are traditionally analyzed separately, and surprisingly little research has been devoted to understanding intertemporal social decision making. People are often confronted with situations that involve a trade-off between immediate and delayed benefits and costs (i.e., intertemporal decisions), and between one's own and others' welfare (i.e., interpersonal decision). In an intertemporal context, activities that yield immediate benefits and whose costs are delayed are intuitively preferred over plans with immediate costs and delayed benefits. However, it is unclear if this preference holds when another person benefits rather than oneself. For instance, are people more generous when they can immediately benefit the receiver compared to situations where their decisions affect others only in the future? Or are they more willing to help someone when the costs for doing so are delayed compared to cases when the costs are immediate?

In **Chapter 4**, *In search of the sunk cost bias*, I present a study where we investigate the sunk-cost bias, which refers to the stronger tendency to continue an endeavor once an investment has been made (Arkes and Blumer, 1985). The sunk cost bias has been put forward as an explanation for why politicians continue over budget public-works (Ross and Staw, 1993), why firms continue to invest in hopeless projects (Arkes and Blumer, 1985), why people stay in failing relationships (Strube, 1988), or why researchers continue less promising projects instead of starting new ones. A rational decision-maker should ignore the initial investment and base the decision on whether to make the additional investment solely on the expected utility of the available options. On the contrary, for an individual that falls prey to the sunk cost bias, past expenses (e.g., money, time, effort, emotions) constrain decision-making in the present even though do not affect the attractiveness of the available options.

We use an experimental design inspired by the classic project continuation example from the psychological literature: an initial investment needs to be made to start a project and to advance it to a second investment stage. The task allows testing several psychological mechanisms that have been proposed as drivers of the sunk cost effect. Specifically, we focus the discussion on the role of responsibility for the initial investment, on the role of being in the loss domain, and on waste aversion. Additionally, after the experiment, participants answered some of the classic survey questions from the sunk cost literature in psychology. The data provide surprising findings that underline the difference between incentivized and hypothetical scenarios in the study of the sunk-cost bias.

In **Chapter 5** I summarize the main findings of this dissertation and in **Chapter 6** I briefly describe how this research contributes to the field of human decision sciences and to society in general.

Chapter 2

Love and retaliation. Distinct effects of prefrontal cortex on second and third-party punishment and reward

Abstract

Punishment and reward are crucial for sustaining cooperation in societies. However, their neural mechanisms are still poorly understood. It is yet unclear whether the same brain networks involved in punishment are engaged in reward, as very little attention has been paid to the latter. In this preregistered study, we use transcranial magnetic stimulation to test the involvement of the right dorsolateral prefrontal cortex (R dlPFC) and medial prefrontal cortex (mPFC) on punishment and reward of greedy and generous monetary transfers. In a within-subjects design, participants received deep active continuous theta burst stimulation (cTBS) to inhibit the R dlPFC and the mPFC in addition to sham stimulation. Participants took part in a novel dictator game as Receiver (second-party) or Observer (third-party) with the opportunity to punish and reward the Proposer. In half of the trials, punishment and reward were costly to participants. Results indicate that the stimulated prefrontal regions play distinct roles in punishment and reward. Stimulation of the R dlPFC and mPFC decreased punishment of greedy allocations in all conditions except when participants were Observers and had no costs to punish. These findings are consistent with a broad perspective of the implication of R dlPFC and mPFC in self-centeredness, which includes both a personal and a monetary component. Following generous offers, a decrease in costly reward is observed, presumably because stimulation of these regions allows selfish economic motivations to overtake norm enforcement or reciprocity motives. In the absence of cost, however, reciprocity drives the decision, resulting in increased reward of generous offers. Although reciprocal fairness is influenced by brain stimulation, individuals' perceptions of fairness and social norms remain unaffected.

This paper is co-authored with Leticia Micheli, Teresa Schuhmann & Arno Riedl. The research presented in this chapter was supported by the grant for primary data collection of the Graduate School of Business and Economics at Maastricht University.

1. Introduction

From lending a hand to assist friends to helping strangers on the street by giving directions or even money, humans are remarkably cooperative towards both kin and non-kin. This is surprising considering that cooperation often involves a tension between own and others' interests and on many occasions not cooperating makes one better off. Reciprocal fairness, which implies reciprocating selfish behavior with punishment and kindness actions with rewards, is essential to discourage defection and to promote cooperation (Egas and Riedl, 2008; Fehr and Gächter, 2002). Research has shown that individuals engage in punishment when they are directly affected by a violation of the fairness norm (Fehr and Gachter, 2000; Gächter et al., 2008) but also when they are an unaffected third-party (Fehr and Fischbacher, 2004a). Despite being much less investigated than punishment, rewards have also been proved to promote cooperation in second (Andreoni et al., 2003; Sefton et al., 2007; Szolnoki and Perc, 2010) and third-party interactions (Almenberg et al., 2010; Nikiforakis and Mitchell, 2014; Sutter et al., 2009). Given the universal human propensity to punish and reward others, several studies have attempted to unravel the neural correlates of these reciprocal behaviors. Despite the accumulating neuroscientific evidence, it is still unclear whether punishment and reward are simply different sides of the same coin and governed by the same brain networks or if they depend on distinct processes.

Neuroscientific studies have repeatedly pointed to the role of the prefrontal cortex in reciprocal fairness. Specifically, the right dorsolateral prefrontal cortex (R dlPFC) has been implicated in a wide variety of higher-level functions, including the control of selfish impulses (Knoch et al., 2009, 2006; Knoch and Fehr, 2007; Van't Wout et al., 2005), the ability to adapt behavior strategically (Ruff et al., 2013; Spitzer et al., 2007; Strang et al., 2015) and the ability to integrate intentionality and harm for the selection of an appropriate reaction (Buckholtz et al., 2008; Buckholtz and Marois, 2012; Buckholtz et al., 2015). The medial prefrontal cortex (mPFC) has also been implicated in reciprocal fairness. In this context, it has been suggested that this region is associated with a positive affect response to fairness (Tabibnia et al., 2008), with mentalizing processes when reacting to unintentional unfairness (Güroğlu et al., 2010) and processing unfairness when it damages the self (Civai et al., 2015; Corradi-Dell'Acqua et al., 2013).

However, the differences in the study designs used in the literature make it difficult to draw comprehensive conclusions about the cognitive mechanisms associated with punishment and reward that generalize beyond the specific conditions tested in these studies. For instance, many studies in the literature focus solely on cases where the person who can reciprocate is the target of someone's actions (e.g., Knoch et al. (2009, 2006)) or where they are unaffected observers (e.g., Buckholtz et al. (2008, 2015)). In the few cases where personal involvement was varied within a study (Civai et al., 2015; Corradi-Dell'Acqua et al., 2013),the cost of reciprocating was also varied, making it hard to disentangle the effect of personal involvement from the effect of costs. In addition, it is still unclear whether R dlPFC and mPFC play a similar role in regulating rewards as they do in punishment, considering that very little attention has been paid to the neural mechanisms of reward in the reciprocal fairness literature (see Knoch et al. (2009) for a notable exception) and we are not aware of any studies that simultaneously investigate the neural mechanisms of punishment and reward at the individual level.

To fill this gap, we present a unified methodological framework where, in a comprehensive within-subjects design using transcranial magnetic stimulation (TMS) to inhibit the functionality of the R dlPFC and mPFC, we provide evidence of the involvement of these prefrontal regions in punishment and reward. Participants took part in a novel monetary allocation game with the opportunity to punish or reward a Proposer after learning how the Proposer divided a monetary amount with the Receiver. During the task, we independently vary personal involvement as each participant receiving TMS takes the role of Receiver (second party) in half of the trials and Observer (third party) in the other half. Also, we vary the cost of punishing and rewarding, such that participants face costly punishment and reward only in half of the trials. Each participant receives stimulation over the R dlPFC, mPFC, and sham stimulation in three separate sessions before completing the monetary allocation game. At the end of each experimental session, participants are also asked to rate the fairness and social appropriateness of Proposer's offers.

This experimental design exhibits several desirable features as we are able to (1) analyze how punishment and reward relate to each other at the behavioral level, (2) discuss whether the different mechanisms associated with the R dlPFC and mPFC extend beyond the contexts in which they have been originally tested, and (3) provide evidence of the differences and similarities between punishment and reward regarding the engagement of the R dlPFC and mPFC. In addition, we provide insights into how punishment and reward are related to motives such as outcome maximization, norm enforcement and reciprocity. Ex-ante hypotheses are derived from literature and pre-registered together with an analysis plan (https://osf.io/pg7yc).

Studies investigating reciprocal fairness have often implicated the R dlPFC in punishing unfair behavior. This region has been shown to be activated when individuals punish non-cooperative behavior (Kodaka et al., 2012), when they interact with non-cooperative partners (Suzuki et al., 2011) and refuse unfair allocations in the Ultimatum Game (Sanfey et al., 2003). Since the dlPFC has been previously associated with *cognitive control* mechanisms, subsequent studies using TMS have proposed that this region is necessary to inhibit prepotent responses to behave selfishly and maximize one's own economic resources while facilitating the implementation of costly punishment in second-party interactions (Knoch et al., 2006; Van't Wout et al., 2005). Consistent with this interpretation, studies have shown that activation in the R dlPFC correlates with the implementation of fair behavior when this is strategically advantageous (Spitzer et al., 2007; Strang et al., 2015).There is also some evidence for the involvement of the R dlPFC in costly rewarding trust in second-party interactions (Van Den Bos et al., 2009). Furthermore, there is evidence that inhibition of this region with TMS decreases the ability to reciprocate trust whenever self-control efforts are necessary to reciprocate in order to build a positive reputation (Knoch et al., 2009). This indicates that the *cognitive control* mechanism of the R dlPFC applies also in the reward domain. This interpretation, however, has been challenged - in studies where reciprocity is absent - by findings showing that inhibition of the dlPFC using cTBS increases offers in a Dictator Game (Christov-Moore et al., 2017) and that cathodal transcranial current stimulation (tDCS) over right lateral prefrontal cortex decreases immediate selfish responses in voluntary transfers, while anodal tDCS leads to a higher degree of immediate selfishness (Ruff et al., 2013).

Importantly, the *cognitive control* hypothesis for the R dlPFC in reciprocal fairness has been exclusively tested in second-party interactions, when it is costly to punish greedy and reward generous behaviors. Therefore, it is still unclear whether this mechanism applies also in other contexts. One important aspect of our study is that we test whether the *cognitive control* mechanism extends also to third-party situations and to contexts where punishing and rewarding are not costly. Building on previous evidence for the involvement of R dlPFC in reciprocal fairness (Knoch et al., 2006), we formulate our first hypothesis, which posits that if the R dlPFC is crucial for overriding economic selfishness impulses, its inhibition with TMS should lead to decrease in punishment and reward only when these are costly and regardless of personal involvement (i.e., second and third-party interactions when punishment and reward are costly). This decrease in punishment and reward should not be observed when punishment and reward are not costly, given that in these conditions economic selfishness could not be a motivation of individuals' behavior. Thus, cognitive control should not be important in facilitating the implementation of reciprocal fairness.

An alternative explanation for the role of R dlPFC in punishing unfair behavior has been proposed by a different stream of literature, which investigates punishment in third-party interactions using hypothetical scenarios instead of economic games with real incentives. Buckholtz et al. (2008) discuss fMRI evidence showing that the same region of the R dlPFC previously implicated in *cognitive control* mechanisms of reciprocal fairness (Knoch et al., 2006) is involved when disinterested third-party participants respond to individuals who are responsible for harming others and make non-costly decisions about appropriate punishments to such transgressions. A follow-up study demonstrated that TMS over the R dlPFC significantly reduced non-costly punishment for wrongful acts in a third-party context (Buckholtz et al., 2015). According to the authors' interpretation of the findings, the underlying process that accounts for the engagement of R dlPFC in third-party punishment is the integration of information about harm and blame and the selection of a specific and appropriate response among a range of response options (Buckholtz and Marois, 2012). This *integration-and-selection* hypothesis implies a general role of R dlPFC in reciprocal fairness, which should apply equally in second and third-party interactions (Buckholtz and Marois, 2012). However, it remains to be explored whether the *integration-and-selection* mechanism generalizes beyond the investigated contexts and applies to second-party interactions and to contexts when punishment is costly. Based on the existing evidence we formulate our second hypothesis: holding responsibility constant, if the R dlPFC is needed to generally determine the appropriate punishment regardless of personal involvement or costs to punish, inhibition of this region with TMS should lead to a decrease in punishment in all conditions compared to sham stimulation (i.e., second and third-party interactions when punishment is costly or not). Importantly, we note that the *integration-and-selection* mechanism has not yet been tested in the reward domain. Here, we speculate that this mechanism should presumably work in the same way in the reward domain. That is, we expect that the R dlPFC is important in integrating information to select an appropriate reward for cooperative and fair behavior. Thus, we hypothesize a decrease in reward in all conditions following inhibition of the R dlPFC if this region is involved in integration-and-selection processing.

Regarding the role of mPFC on reciprocal fairness, existing functional imaging evidence indicates that the mPFC is significantly more engaged in second party that in third party interactions. Specifically, Corradi-Dell'Acqua et al. (2013) shows that the anterior mPFC is more active when a participant is confronted with unfair offers in the Ultimatum Game, compared to a situation where, as a disinterested third party, the participant observes another individual receiving an unfair offer and decides on the behalf of this individual whether to accept or reject the offer. A follow-up study later confirmed the role of the anterior mPFC in second-party punishment decisions by applying cathodal stimulation on the mPFC, which resulted in decreased rejection rates of unfair offers in the second party, but not in the third-party condition (Civai et al., 2015). This engagement of mPFC in reciprocal fairness is consistent with neuroscientific evidence which has repeatedly implicated cortical midline structures -including the mPFC- in processing self-referential stimuli, for example during self-reflection (Johnson et al., 2002; Kelley et al., 2002) and evaluative judgments that depend on one's internalized values (Zysset et al., 2002) (for a review, see Northoff and Bermpohl (2004); Northoff et al. (2006)). According to this *self-referential* interpretation, in the context of reciprocal fairness, the mPFC is especially recruited when punishing unfairness that is directed towards oneself. However, studies investigating the role of mPFC on reciprocal fairness have only compared costly second-party punishment with non-costly third-party punishment. The fact that both costs and personal involvement were simultaneously varied leaves room for a *cognitive control* interpretation of results, according to which the mPFC could be involved in overriding economic selfish impulses in favor of implementing fair behavior (Civai et al., 2015). This would also explain the observed decrease in punishment following inhibition of the mPFC only in second-party interactions, where punishing was costly (Civai et al., 2015). Based on this evidence, we formulate our third hypothesis: if the anterior mPFC is important for *self-referential* thinking during reciprocal fairness, its inhibition should result in a decrease of punishment in second-party conditions compared to sham stimulation, regardless of the cost of punishing. Such an effect should not be observed in third party conditions. We note that the *self-referential* mechanism has not been tested in the reward domain. We propose that a similar process as in punishment could take place when people decide whether and how to reward fair behavior. Alternatively, and according to our fourth hypothesis: if the mPFC is involved in cognitive control mechanisms which inhibit economic selfishness to enforce implementation of fairness, as proposed by Civai et al. (2015), inhibition of the mPFC should lead to reduced punishment and reward in conditions where it is costly, regardless of personal involvement.

Lastly, previous brain stimulation literature has shown that inhibition of R dlPFC affects reciprocal fairness without altering judgments of what is fair (Knoch et al., 2006; Strang et al., 2015). We assume that, just like the R dlPFC, the mPFC is important for implementing fairness behavior but is not involved in assessing the fairness of a given situation. Our fifth hypothesis then posits that inhibition of R dlPFC and mPFC does not affect fairness and social appropriateness judgments. Table 2.1 summarizes the hypotheses formulated in the present study.

Table 2.1: **Summary of hypotheses.**

Hypotheses	Explanation
H1: Cognitive control (R dlPFC)	Inhibition of the R dlPFC should lead to decrease in punishment and reward only when these are costly and regardless of personal involvement (i.e., second and third-party interactions when punishment and reward are costly). This decrease should not be observed when punishment and reward are not costly.
H2: Integration and selection (R dlPFC)	Inhibition of the R dlPFC should lead to a decrease in punishment and reward in all conditions compared to sham stimulation (i.e., second and third-party interactions when punishment is costly or not).
H3: Self-referential (mPFC)	Inhibition of the mPFC should result in a decrease of punishment and reward in second-party conditions compared to sham stimulation, regardless of the cost of punishing. Such an effect should not be observed in third party conditions.
H4: Cognitive control (mPFC)	Inhibition of the mPFC should lead to decrease in punishment and reward only when these are costly and regardless of personal involvement (i.e., second and third-party interactions when punishment and reward are costly). This decrease should not be observed when punishment and reward are not costly.
H5: Fairness and social appropriateness (R dlPFC and mPFC)	Inhibition of R dlPFC and mPFC does not affect fairness and social appropriateness judgments.

2. Methods

2.A. Participants

Thirty-three participants (24 women, mean age= 21.9 years, min = 18 years, max = 33 years, SD=3.4 years, 29 right-handed) took part in the experiment. The study was approved by the local ethical review committee of Psychology and Neuroscience at Maastricht University (ERCPN-192_19_04_2018). Participants were screened to ensure they met the requirements to receive transcranial magnetic stimulation (Rossi et al., 2009) as well as to guarantee they were never exposed to experiments involving deception. Only participants with normal or corrected-to-normal vision and without any history of neurological or psychiatric disorders were included.

2.B. Procedures

Participants attended three different sessions, in which they received, in a counterbalanced order, TMS aiming to inhibit the functionality of the R dlPFC, anterior mPFC or sham stimulation (control). For each participant, sessions were scheduled at least one week apart from each other. Out of a total of 96 sessions, 12 were conducted with less time between sessions due to scheduling restrictions but were always minimally separated by 4 days. In each session, participants filled in a TMS screening form and provided written consent. At the start of the first session, participants received general information about TMS and their active motor threshold was determined (see “TMS Procedures” for more details). Prior to brain stimulation, they received written instructions about the experiment task and answered comprehension questions. After that, an electroencephalogram (EEG) cap indicating the electrode positions of the

International 10-20 EEG system was placed on participants' heads to determine the site of stimulation. Cap sizes were chosen based on participants' head size and were adjusted so that the Cz position would be placed in the middle point between right and left preauricular points and between the nasion and the inion. TMS was then applied and participants started the task immediately after receiving brain stimulation.

Each experimental session lasted on average 75 minutes and participants received a participation fee of €7.50 for each one. In addition, decisions in the experiment were incentivized and participants could earn extra money (between €0 and €25.30) in each session depending on their decisions and on the decisions of the individuals they were matched with. On average, participants earned €25.80 (SD= € 6.60) per experimental session.

2.C. TMS Procedures

A MagVenture MagPro X100 stimulator equipped with the high-performance cooling system (MagVenture, Farum, Denmark) and a figure-of-eight TMS coil (Cool- D-B80; inner diameter = 67 mm, outer diameter = 95 mm) were used to deliver continuous theta-burst stimulation (cTBS). This coil was selected given its balance between depth of penetration and focality. Because of its geometry, it is suitable for reaching deep regions in the brain (i.e., the mPFC in our study), while still being able to produce a focal stimulation (Deng et al., 2013). The protocol consisted of 50Hz triplets delivered 5 times a second for 40 seconds (600 pulses in total) (Huang et al., 2005). This TMS protocol has been shown to produce consistent and rapid electrophysiological and behavioral changes that last for approximately 60 minutes after stimulation, which is considerably longer than other more traditional stimulation protocols (Huang et al., 2005).

The stimulation intensity was 90% of the individual active Motor Threshold (aMT), defined as the lowest machine output intensity able to induce a visible twitch in the stretched out left index finger for 50% of the pulses. The aMT was determined in the first session and used for the other three sessions (mean stimulator output= 29.1% of maximum stimulator output, SD = 4.29; mean realized output = 42 $A/\mu s$). The same protocol was applied to sham stimulation, but the coil was tilted at 180 degrees.

The stimulation sites were localized using the international 10-20 EEG system, which has been shown to provide a reliable cortical positioning for large scale cortical areas such as the ones our study targets (Herwig et al., 2003). In all three sessions, participants wore a cap with EEG electrodes, so that stimulation sites could be determined. Based on the Talairach coordinates provided by Knoch et al. (2006) and by Buckholtz et al. (2008) for the R dlPFC (x=39, y=37, z=22), we placed the center of the TMS coil on the electrode site F4 while holding the handle of the coil parallel to the cortical midline. Likewise, based on the MNI coordinates reported by Corradi-Dell'Acqua et al. (2013) for the anterior mPFC (x=0, y=58, z=8), this brain region was stimulated by placing the coil between Fp1 and Fp2 while holding the handle of the coil at a 45 degrees angle from the cortical midline. For sham stimulation, the coil was placed in the middle point between the stimulation sites of the R dlPFC and mPFC.

2.D. Task: Dictator game with punishment and reward

In each brain stimulation session, participants receiving TMS (henceforth TMS participants) made decisions in 140 trials. In our task, a Dictator Game with punishment and reward, three individuals are randomly matched and are assigned three different roles: Proposer, Receiver, and Observer (task adapted from Spitzer et al. (2007)). TMS participants always made decisions in the role of Receiver and Observer, and never in the role of Proposer. In each trial of the game, the Proposer is endowed with 240 experimental currency units (ECUs), the Receiver is not endowed with any ECUs, and the Observer is endowed with 200 ECUs. All Participants are informed that at the end of the experiment ECUs are converted to Euro at a rate of €0.10 per ECU. In each trial, TMS participants are confronted with a Proposer's decision of how to share his endowment with the Receiver. Proposers could choose between different allocation bundles which range from very greedy to very generous. Specifically, proposers could choose from the offers, 200-40, 160-80, 120-120, 80-160, and 40-200, where the first term refers to what the Proposer keeps for himself or herself and the second term refers to what is allocated to the Receiver (see section "Proposers" for a more detailed description of the possible allocations available to Proposers).

Depending on the experimental condition, either the Receiver or the Observer can punish or reward by decreasing or increasing the Proposer's payoff after learning the allocation decision made by the Proposer. A 2x2 within-subjects experiment design was implemented in which both personal involvement (Receiver vs Observer) and the cost to punish or reward the Proposer (Cost vs. No_cost) vary (see Figure 2.1). In both the Receiver with Cost (Rec_C) and Receiver with No Cost (Rec_NC) conditions, the TMS participant makes decisions in the role of the Receiver and is thus directly affected by the Proposer's choice. In the Rec_C condition, each ECU invested in punishment and reward, respectively, reduces and increases the payoff of the Proposer by 5 ECUs, while decreasing the payoff of the Receiver by 1 ECU. Our design ensured that irrespective of the Proposer's allocation decision, the Receiver always has enough ECUs to invest in punishing or rewarding the Proposer. In the Rec_NC condition, punishing or rewarding the Proposer is not costly to the TMS participant. That is, every ECU invested by the Receiver in Rec_NC reduces or increases the Proposers' payoff by 5 ECUs but does not reduce the Receiver's payoff.

In the Observer with Cost (Obs_C) and Observer with No Cost (Obs_NC) conditions, the TMS participant makes decisions in the role of the Observer. That is, the TMS participant observes the allocation decision of the Proposer regarding the Receiver but is otherwise not affected by it. After learning the allocation, the TMS participant can use the endowment of 200 ECUs to either reward or punish the Proposer. Equivalent to the Receiver conditions, in Obs_C an ECU used for rewarding and punishing the Proposer respectively increases and reduces the Proposer's payoff by 5 ECU and the Observer's payoff by 1 ECU. In the Obs_NC condition, the effect was the same for the Proposer but there are no costs for the Observer.

In all four conditions (i.e., Rec_C, Rec_NC, Obs_C, and Obs_NC), TMS participants are informed they can invest up to 40 ECUs to reward or punish the Proposer. This ensures that in the case of the most greedy allocation by the

Proposer, the TMS participant in the Rec_C condition has sufficient ECUs to decrease the Proposer's payoff to zero. Moreover, TMS participants are informed they cannot decrease Proposers' payoff below 0 or increase Proposers' payoff above 240 ECUs. These rules ensure that TMS participants could never incur losses nor cause losses to Proposers.

Within each condition, the TMS participant and the Proposer are the only individuals who are aware of the decision taken by the TMS participant to punish or reward the Proposer. This means that in the Rec_C and Rec_NC conditions, the TMS participant's decision as Responder to punish or reward the Proposer cannot be seen by the Observer. Likewise, in the Obs_C and Obs_NC conditions, the TMS participant's decision as Observer to punish or reward the Proposer cannot be seen by the Receiver. This ensures that the decisions to punish and reward made by TMS participants are not influenced by reputation or image concerns that could ensue from being observed. Moreover, our set-up implies that in each condition there is only one person who can modify the Proposers' payoff and one person who can only passively watch the behavior of the Proposer. Hence, information is kept symmetric in all conditions.

In each brain stimulation session, TMS participants face the four different conditions in a randomized block design. In each block, TMS participants are confronted with 35 different proposer's allocations presented in random order, which vary from very greedy to very generous. After the 35 allocations for a given condition are shown to participants, a new block starts featuring a new condition. During this task, facial expressions of the TMS participants were recorded with a webcam. Results of facial expression analysis are not reported here.

At the end of each experimental session, one trial was selected randomly for payment and all three participants in that trial were paid accordingly. Session earnings and total earnings were revealed to the TMS participants only at the end of the last experimental session to avoid that earnings in a session would influence their decisions in the subsequent sessions. All earnings, including participation fees, were transferred to TMS participants' bank accounts within one week from their last experimental session. For a detailed explanation of how the payoffs of each participant were calculated, see the Appendix.

Although every trial required the participation of three individuals, only the participant undergoing brain stimulation was present in the laboratory. Proposers, as well as passive participants (i.e., Observers in Rec_C and Rec_NC and Receivers in the Obs_C and Obs_NC), were real participants, but they were not present during the brain stimulation sessions. Nevertheless, the rules of the task were common knowledge to all of them. The participation of Proposers and passive participants in the experiment is explained in detail below.

Proposers

The present experiment does not involve deception, which means TMS participants faced real Proposers' allocations and were matched with a different Proposer in each trial. Proposers' allocations were collected before the brain stimulation experiment via the online platform Prolific (www.prolific.co). Prior to making any decisions, they were informed their choices could be selected and used in a later study and that the rules of this later study would apply to them. Proposers were not informed that the future study would involve brain stimulation. Proposers (N=211) received £0.75 for completing the online study, where they read instructions about the dictator game with punishment and reward and were presented with the four different conditions of the experiment. In each condition, Proposers were endowed with 240 ECUs and were asked to indicate how they would like to allocate the ECUs between themselves and a Receiver. Proposers knew they could be contacted later to receive their earnings according to their allocation decisions and the received reward or punishment in the selected trial. Proposers selected one of five possible allocation bundles (see Table 2.2). For each of these bundles, they were truthfully told that the researchers could select any allocation within the range of ±3 ECUs from their original choice. For instance, if a Proposer chose the allocation bundle 200-40, the Receiver could be shown an allocation between 197-43 and 203-37. This procedure allowed us to collect allocations that were real and yet showed some variation, as to keep the task engaging to participants undergoing brain stimulation. Importantly, TMS participants were truthfully informed about the options the Proposers had and about the way researchers could select their choices. After the conclusion of the experiment, Proposers were contacted and paid according to their final payoff in the trial selected for payment in the brain stimulation study. All payments to Proposers were transferred via the online platform Prolific.

Table 2.2: **Monetary allocations available to Proposers.**

Allocation Bundle	Classification	Proposer payoff	Receiver payoff
1	Very Greedy	200±3	40±3[1]
2	Greedy	160±3	80±3
3	Equal	120±3	120±3
4	Generous	80±3	160±3
5	Very Generous	40±3	200±3

Note: Monetary allocations are here for convenience labeled ranging from "Very greedy" to "Very generous". These labels were not used in the experiment. Proposers chose an allocation bundle per condition, and a specific allocation within that bundle (±3 ECUs) could be presented to a TMS participant during the brain stimulation study. This was common knowledge to all the participants.

Passive participants

Observers in the Rec_C and Rec_NC conditions and Receivers in the Obs_C and Obs_NC conditions are passive participants as they could not modify the payoff of any other individual in the experiment. Passive participants (N=96) were contacted via the ORSEE online recruitment system for economic experiments (Greiner, 2015) of the Behavioral and Economics Laboratory (BEELab) at Maastricht University and asked to participate in a study. They received instructions about the dictator game with punishment and reward and its four conditions. Thereafter, they received an email with information about the Proposers' allocation decision in the trial selected for payment and their own earnings in that specific trial. All payments were transferred via bank transactions.

[1]In case of receiving the most selfish allocation (i.e., 37 ECUs), the Receiver in the 2p_c condition could use all his ECUS to bring the Proposer's payoff close to zero.

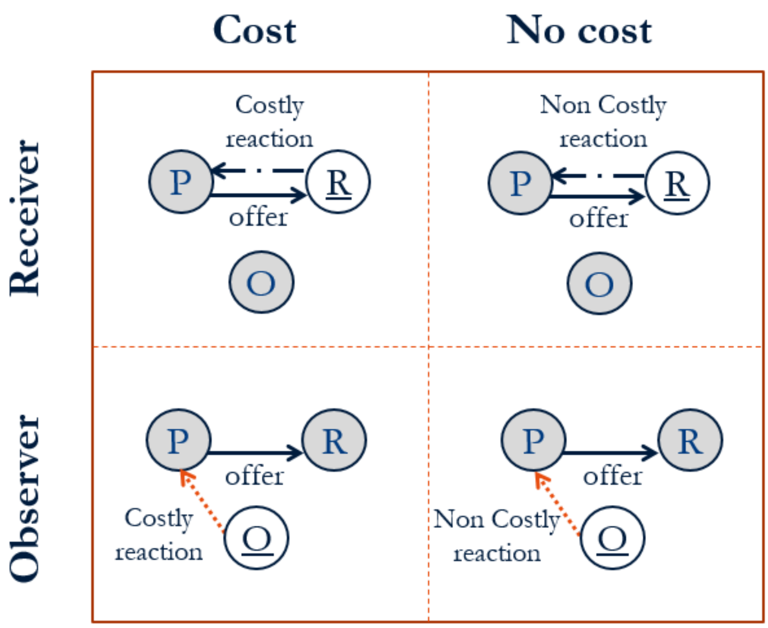


Figure 2.1: **Dictator game with Punishment and Reward**

Note: In each trial, three players are involved: a Proposer (P), a Receiver (R) and an Observer (O). The TMS participant (indicated by white circles) makes decisions in the roles of the Receiver (top panels) and in the role of the Observer (bottom panels). The two panels on the left refer to the conditions where it is costly for the TMS participant (Receivers in the Rec_C condition and Observers in the Obs_C condition) to modify the Proposer's payoff. The two panels on the right refer to non-costly conditions, where the TMS participant (Receivers in the Rec_C condition and Observers in the Obs_C condition) can modify the Proposer's payoff without bearing any costs.

2.5. Fairness and appropriateness of social behavior

Immediately after the Dictator game with Punishment and Reward, TMS participants completed two tasks to test whether brain stimulation affected the perceptions of what constitutes fair and socially appropriate behavior, respectively. Fairness ratings measure the perceived fairness of the different proposals and social appropriateness judgments measure descriptive social norms. Fairness and appropriateness ratings were presented in a counterbalanced order and were completed within the range of the TMS intervention (i.e., 25 to 30 minutes after brain stimulation).

Fairness ratings

For each of the four conditions, TMS participants were asked to judge the fairness of each of the bundle allocations available to Proposers (40-200; 80-160; 120-120; 160-80; 200-40) on a 7-point Likert scale ranging from "very unfair" to "very fair".

Appropriateness ratings

TMS participants were asked for the social appropriateness of the five possible Proposer's bundle allocations in each condition using the incentive-compatible norm elicitation task introduced by Krupka and Weber (2013). Social appropriateness was described to participants as the "behavior that most people agree is the correct or ethical thing to do". Appropriateness was rated on a 4-point Likert scale that ranged from "Very socially inappropriate" to "Very socially appropriate". At the end of each brain stimulation session, one of the Proposer's allocations was randomly selected. The appropriateness rating of the TMS participant for the selected allocation was then compared to the most frequent answer for the same allocation provided by another group of individuals, whose answers were collected in a separate online experiment that took place before the brain stimulation experiment. Collecting answers from a separate group of participants, which were not exposed to brain stimulation, is important as it allows avoiding any deception, ensures the more frequent social appropriateness answer is somewhat representative of the general population beliefs and avoids that these answers are biased by the effects of TMS. The participants in this online experiment received information about the four different conditions and were asked to rate, for each condition, the social appropriateness of each of the five allocation bundles that the Proposer could choose. Their ratings were incentivized with a €5 payment if their appropriateness judgment matched the most frequent answer in their group.

Each TMS participant received €5 if his or her social appropriateness judgment matched the most frequent answer in the separate online experiment. This method is commonly used to measure social norm perception in an incentive-compatible way (Banerjee, 2016; Exley et al., 2020; Gneezy et al., 2014; Thomsson and Vostroknutov, 2017). TMS participants were only informed about the earnings of the appropriateness task at the end of the last experimental session.

2.6. Data Analysis

As the neural mechanisms of reward have never been explored in relation to those of punishment, we investigate these two domains separately to be able to identify potential similarities and dissimilarities between them. For punishment, we focus on trials where TMS participants received Very Greedy (200-40) and Greedy allocations (160-80). Punishment captures almost all responses to these allocations, as counterintuitive reward of Greedy and Very Greedy allocations was rare (less than 5.2% of the cases). For reward, we consider trials in which TMS participants received Equal (120-120), Generous (80-160) and Very Generous allocations (40-200). As before, reward captures almost all responses to these allocations. Counterintuitive punishment of Equal, Generous and Very Generous allocations was also negligible (less than 6.4% of all cases). In both the punishment and reward domains, the dependent variable of interest is the impact of punishment and reward on

Proposer's payoff. For ease of presentation, we code a decrease in the Proposers' payoff due to punishment to positive values. Hence, in both the punishment and reward domains, a value of zero means no punishment or reward and a value of +200 means maximum punishment or reward. Our within-subjects design, where the same participants repeatedly made decisions in the different conditions, provides a panel data set. In addition, the dependent variable is censored at 0 and 200 ECUs. Therefore, in our analysis, we use random-effects Tobit regressions for panel data that are left- and right-censored with standard errors clustered at the participant level. In all regression models, we control for gender, fairness (1-7) and appropriateness ratings (1-4), trial order within an experimental session (1-140), and session number (1-3).

We start by reporting sham stimulation results. We first describe how fair and socially appropriate TMS participants considered the Proposers' offers. We then report whether, without an effective brain stimulation, it matters that punishment and reward are costly or not and whether there is a difference in responses between Receiver and Observer conditions. In the regression model, dummy predictors are used for the roles of Receiver or Observer and whether punishment or reward was costly or not.

We then turn our focus to the effect of stimulation of the R dlPFC and mPFC on TMS participants' responses compared to sham. First, we conduct a Kruskal-Wallis equality-of-populations rank test to investigate whether brain stimulation affects participants' fairness judgments and socially appropriateness ratings of the different experimental allocations. Second, to examine the effect of brain stimulation on punishment or reward, we include in the regression dummy variables coding stimulation (R dlPFC, and mPFC) considering sham stimulation as the baseline category, the two-way interactions between stimulation and cost of punishment or reward and stimulation and personal involvement, and a three-way interaction between stimulation, cost of punishment or reward and personal involvement. For clarity of exposition, we present the marginal effects of brain stimulation relative to sham in the four different conditions. Marginal effects are computed at the censored expected value. All p-values associated with the comparison between the effect of brain stimulation (i.e., R dlPFC and mPFC) and sham in the four experimental conditions survive multiple comparisons correction using the False Discovery Rate (FDR) method (Benjamini and Hochberg, 1995) at 95% confidence level, unless otherwise stated. All analyses reported were conducted with the statistical software Stata 15.

Identifying Weak Reciprocators

We hypothesize that, depending on the functional role of the R dlPFC and mPFC, inhibition of these areas with TMS results in reduced punishment and reward. To test these hypotheses, it is crucial that TMS participants punish and reward Proposers following sham stimulation, such that there is room to observe a reduction of punishment and reward between sham and the other brain stimulation conditions. We identified 8 participants who, across all experimental conditions in sham, decided to not punish or reward Proposers in more than 50% of all trials in sham and for whom it would be hard to test the hypothesized effects of TMS stimulation. We refer to these participants as Weak Reciprocators and we do not include them in the main analysis reported in the paper.

This leaves us with a final sample for our main analysis of 25 participants (mean age = 22.28, SD= 3.64, 19 women) who each reciprocated Proposers' offers in at least 50% of the trials and participated in the three brain stimulation sessions[2]. Importantly, we note that differences in results obtained with this sample and the full sample (which included also the Weak Reciprocators) are not extensive, and mostly related to the punishment domain. However, as this exclusion criteria was not included in our pre-registered analysis plan, results for the whole sample of participants are also presented (see the Appendix).

3. Results

Fairness and social appropriateness following sham stimulation

As fairness and social appropriateness ratings might be related to participants' punishment and reward behavior, we start by examining how TMS participants judge the Proposers' allocations following sham stimulation. As Figure 2.2 shows, fairness and social appropriateness ratings vary depending on Proposers' offers during sham stimulation. Very Greedy (200-40) and Greedy (160-80) allocations are in general rated as unfair and socially inappropriate, whereas the equal split (120-120) is the bundle of allocations considered by participants as being the fairest and most socially appropriate. Interestingly, Generous (80-160) and Very Generous (40-200) allocations are perceived as less fair and less socially appropriate than the equal split. There are no differences in fairness and social appropriateness ratings between the four experimental conditions (Kruskal-Wallis equality-of-populations rank test, p= .998 and p=.979 respectively).

[2]One participant did not complete all three brain stimulation sessions. Excluding this participant from the main analysis only changes the significance of results in one specific condition, which will be mentioned when reporting the respective results.

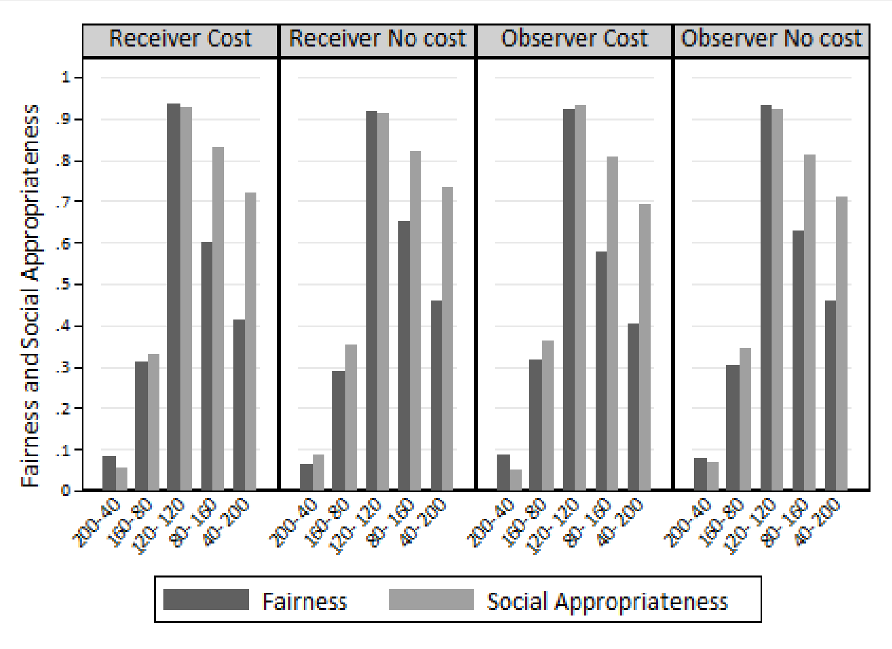


Figure 2.2: **Normalized fairness and social appropriateness ratings following sham stimulation.**

Note: Fairness and social appropriateness ratings are shown per monetary allocation and per condition in the sham stimulation. As fairness and social appropriateness questions were answered using different scales, we normalized both ratings on a scale ranging from 0 (Very unfair/Very Social inappropriate) to 1 (Very fair /Very socially appropriate) to facilitate comparison.

3.1. Punishment following sham stimulation

In Figure 2.3, the light grey lines show the pattern of participants' punishment behavior following the sham stimulation for the different cost and personal involvement conditions. After receiving Greedy and Very Greedy offers in the sham stimulation, participants punish significantly more in the Receiver than in the Observer conditions ($\beta = 7.92, p < 0.05$) and punishment is significantly lower when punishing is costly than when it is not costly ($\beta = -74.0, p < 0.001$). The interaction term between the cost of punishment and personal involvement represents the difference of the effect of cost of punishing for Observers compared to Receivers ([Obs_C – Obs_NC] – [Rec_C – Rec_NC]). Results indicate that the difference in punishment between costly and non-costly punishment is significantly larger when the TMS participant is the Receiver than when the TMS participant is in the role of the Observer (β=-31.9, p < 0.001).

Judgments of fairness and social appropriateness influence participants' decisions to decrease Proposers' payoff during the sham treatment. Specifically, the more participants consider Proposers' offers to be unfair and socially inappropriate, the more they punish Proposers (β_{fair} =-7.10, p <0.01; β_{appro} = -22.66, p <0.001).

3.2. *Reward following sham stimulation*

In Figure 2.3, the black lines show the pattern of participants' reward behavior in the sham stimulation for the different cost and personal involvement conditions. The cost of reward is an important predictor of the decision to reward Equal, Generous and Very Generous offers under sham stimulation. Participants reward significantly less when it is costly than when it is not costly ($\beta = -60.44, p < 0.001$). In contrast to the punishment domain, personal involvement does not affect how much participants reward Proposers ($\beta = 5.11, p = 0.123$). Also, the interaction between the cost of norm enforcement and personal involvement is not significant ($\beta = -4.23, p = 0.52$). Interestingly, we find a negative relationship between fairness and altruistic reward ($\beta = -22.05, p < 0.001$), indicating that overly generous offers are rewarded more although they are viewed as less fair than the equal split. Social appropriateness judgments of allocations have no effect on participants' decisions to reward Proposers ($\beta = 1.87, p = 0.53$).

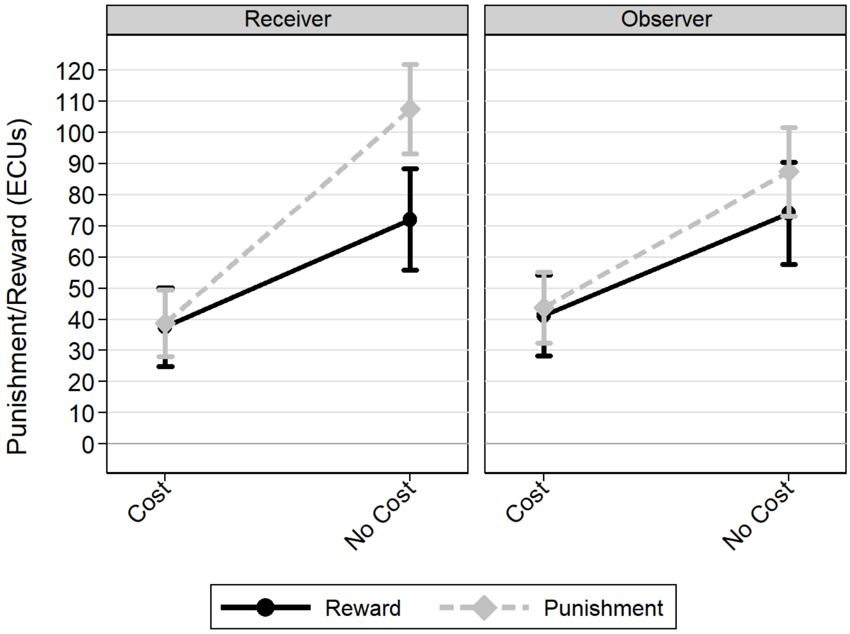


Figure 2.3: **Altruistic punishment and Reward in sham stimulation.**

The figure shows the estimated coefficients for punishment and reward expressed as the absolute change on the Proposer's payoff (i.e., 0 = no punishment/reward; +200 = maximum punishment/reward) in different conditions, after sham stimulation. The circles and diamond markers represent the point estimates for reward and punishment, respectively. The bars indicate the 95% confidence intervals.

3.3. The effect of TMS on fairness and social appropriateness

Figure 2.4 shows the fairness and appropriateness ratings of Proposers' allocations for sham and for the brain stimulations over R dlPFC and mPFC. Following any brain stimulation (i.e., dlPFC, mPFC) we do not observe any differences in fairness and social appropriateness ratings between the four experimental conditions (Kruskal-Wallis equality-of-populations rank test: $p > 0.941$ for both fairness and social appropriateness). We therefore aggregate the ratings within a brain stimulation session and we find that stimulation of R dlPFC or mPFC did not alter fairness and social appropriateness judgments compared to sham (Kruskal-Wallis equality-of-populations rank test: Fairness $p = 0.222$; Appropriateness $p = 0.249$). Hence, any difference in punishment and reward following brain stimulation compared to sham cannot be attributed to changes in TMS participants' fairness or socially appropriateness judgments.

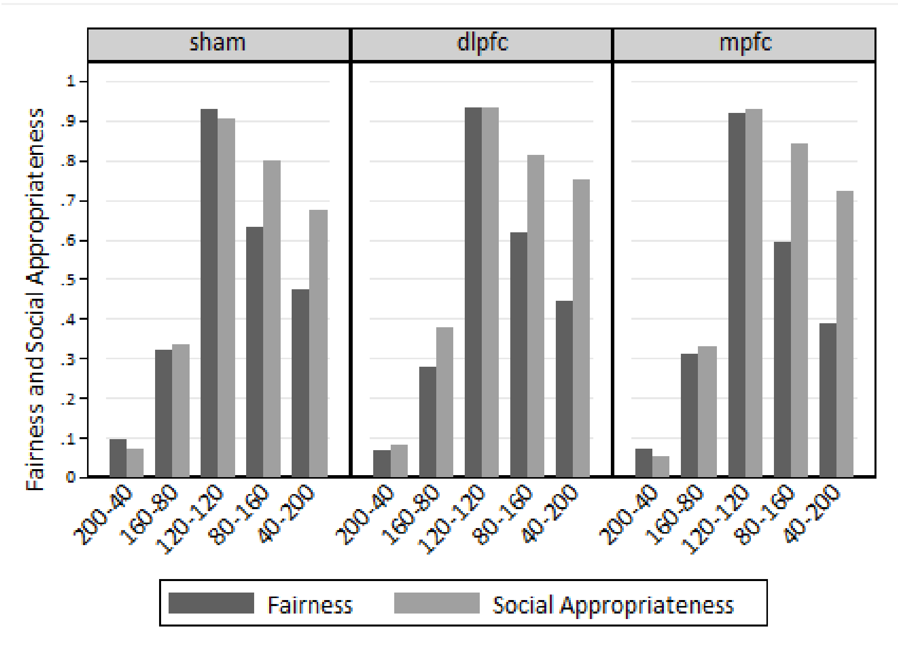


Figure 2.4: **Effect of brain stimulation on fairness and appropriateness ratings.**

Note: Fairness and social appropriateness ratings are shown per stimulation and allocation bundle offered by the Proposer. The ratings shown here are normalized on a scale ranging from 0 (Very unfair/Very Social inappropriate) to 1 (Very fair/Very socially appropriate) to facilitate comparison.

3.4. *TMS effect on punishment*

Figure 2.5 depicts the effect of brain stimulation over R dlPFC and mPFC on punishment relative to sham, for both roles (Receiver and Observer) and both cost conditions (Cost and No Cost).

R dlPFC stimulation

Compared to sham, stimulation of the R dlPFC decreases costly and non-costly punishment of Greedy and Very Greedy offers when TMS participants are in the role of Receivers (Rec_C: t=-9.38, p = 0.001; Rec_NC: t=8.37, p = 0.04), and costly punishment when TMS participants are in the role of Observers (Obs_C: t=-8.15, p = 0.009). Stimulation of the R dlPFC does not affect non-costly punishment for Observers (Obs_NC: t=0.19, p = 0.96).

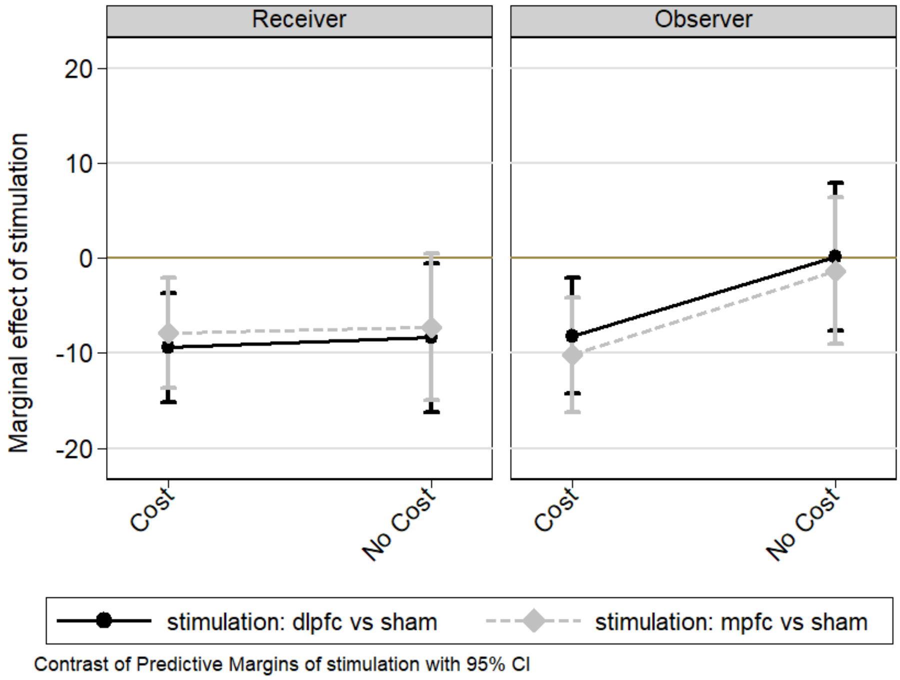


Figure 2.5: **Effect of TMS stimulations on punishment.**

Note: The figure summarizes the marginal effect of R dlPFC and mPFC stimulation relative to sham stimulation on punishment of Greedy and Very Greedy offers. The horizontal solid line at zero indicates the point estimate for the sham stimulation in the different conditions. The black circle and the grey diamond markers represent the marginal effects of R dlPFC and mPFC stimulation compared to sham, respectively. The bars indicate the 95% confidence intervals based on the regression analysis.

mPFC stimulation

Compared to sham, stimulation of the mPFC decreases costly punishment of Greedy and Very Greedy allocations for both Receiver and Observer roles (Rec_C t=-7.87, p = 0.007; Obs_C: t=-10.18, p < 0.001). When punishment is not costly, a decrease in punishment is observed, at a marginal significance level, for TMS participants in the role of the Receiver (Rec_NC: t=-7.26, p = 0.06, FDR corrected <.10)[3], but not for TMS participants in the role of the Observer (Obs_NC: t=-1.31, p = 0.73).

3.5. *TMS effect on reward*

Figure 2.6 shows the effect of brain stimulation over R dlPFC and mPFC on reward relative to sham, for both roles (Receiver and Observer) and both cost conditions (Cost and No Cost).

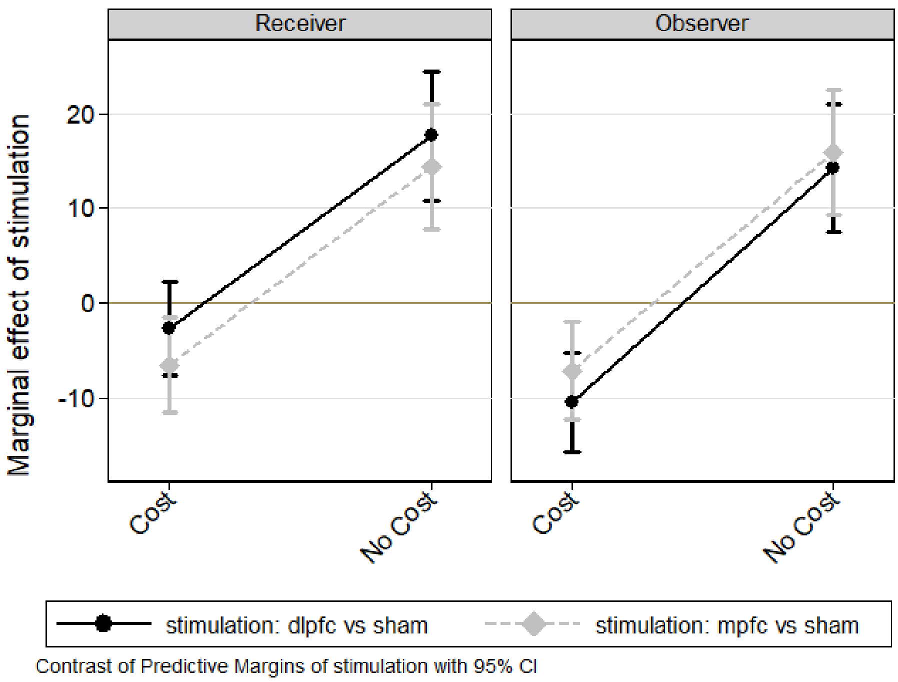


Figure 2.6: **Effect of TMS stimulations on reward.**

Note: The figure summarizes the marginal effect of R dlPFC and mPFC stimulation versus sham stimulation on reward of Equal, Generous and Very Generous offers. The horizontal solid line at zero indicates the point estimate for sham treatment in the different conditions. The black circle and the grey diamond markers represent the marginal effects of R dlPFC and mPFC stimulation compared to sham, respectively. The bars indicate the 95% confidence intervals based on the regression analysis.

[3]We note that when excluding the participant who did not participate in all the three different experimental sessions, we observe that following mPFC stimulation the decrease of punishment in the Rec_NC condition becomes significant t=-8.72, p = 0.03). The other results reported in the paper remain the same when excluding this participant.

R dlPFC stimulation

Compared to sham, stimulation of the R dlPFC decreases costly rewards of Equal, Generous and Very Generous allocations when TMS participants are in the role of the Observer (Obs_C: t=-10.45, $p < 0.001$), but not when TMS participants are in the role of the Receiver (Rec_C: t=-2.67, $p = 0.29$). In stark contrast, stimulation of the R dlPFC strongly increases non-costly reward relative to sham, for both Receiver and Observer roles (Rec_NC: t=17.67, $p < 0.001$; Obs_NC: t=14.27, $p < 0.001$).

mPFC stimulation

Compared to sham, stimulation of the mPFC decreases costly altruistic reward of Equal, Generous and Very Generous allocations, for both Receiver and Observer roles (Rec_C: t=-6.54, $p = 0.01$; Obs_C: t=-7.14, $p = 0.007$). In stark contrast again, stimulation of the mPFC increases non-costly rewards relative to sham, irrespective of the role of the TMS participant (Rec_NC: t=14.43, $p < 0.001$; Obs_NC: t=15.93, $p < 0.001$).

4. Discussion and conclusion

We present a comprehensive test of the involvement of the R dlPFC and mPFC in punishment and reward at the individual level (within-participants design). We independently vary personal involvement (second vs. third party) and the cost of punishment and reward (cost vs. no cost) in an attempt to test, in diverse social contexts, the different explanations that have been given to the role of the R dlPFC and mPFC in reciprocal fairness. Furthermore, we provide a direct comparison of the involvement of these regions in punishing greedy and rewarding generous behavior.

We discuss unexpected results which only partly confirm recently proposed mechanisms for the engagement of these two prefrontal regions in punishment and reward. Our findings show that stimulation of the R dlPFC and the mPFC reduces punishment of unfair actions in conditions where punishing is costly to the punisher and when unfairness is targeted at oneself. A different pattern of results emerges in the reward domain, where stimulation of the R dlPFC and mPFC leads to a decrease in non-costly reward and an increase in costly reward. Below we discuss the relevance of these findings in relation to the existent literature and potential psychological mechanisms that might be associated with reciprocal fairness.

The R dlPFC and mPFC on punishment

Compared to sham, we found that disruption of the R dlPFC decreases costly punishment irrespective of personal involvement. This result is consistent with a *cognitive control* mechanism, suggesting that R dlPFC is needed to override economic selfish impulses when reciprocal fairness is costly (Knoch et al., 2006; Van't Wout et al., 2005). Previous studies have proposed this mechanism in the context of second-party interactions. Here, we replicate these previous findings and provide evidence that disruption of R dlPFC reduces costly punishment also when participants are in the role of unaffected Observers. That is, in situations where punishment is costly, but in which the unkind actions are not directed to oneself.

In addition to reducing costly punishment, we observed that stimulation of the R dlPFC reduces also non-costly punishment when unfair actions are directed to oneself. The only condition where no effect is observed is when the uninvolved Observer has no costs to punish the unfair behavior of others. Because we do not observe a decrease in punishment in all conditions, we argue that our results are not consistent with the *integration-and-selection* mechanism, according to which we hypothesized the R dlPFC would be involved in selecting the appropriate reaction to an unfair offer regardless of personal involvement and cost.

Regarding the mPFC, we find that stimulation of this region leads to decreased costly punishment compared to sham. A marginally significant decrease was also observed in non-costly punishment when the Receiver was the target of the unfairness. Our findings replicate those of Corradi-Dell'Acqua et al. (2013) and Civai et al. (2015), as disruption of the mPFC significantly decreases costly punishment for Receivers compared to sham without affecting non-costly punishment for Observers. However, as mPFC stimulation also leads to a reduction of costly punishment when being an unaffected Observer, our findings do not exclusively support the *self-referential* mechanism but suggest the mPFC is also involved in a *cognitive-control* mechanism.

Altogether, given the similar pattern of results obtained following stimulation of the R dlPFC and mPFC, we suggest that these two regions respond both to the economic costs of punishing and to the target of the unfair action. We believe our results point to the need of revisiting the *cognitive control* and *self-referential* mechanisms in an attempt to unify these two accounts. Our results are consistent with a broad perspective of the implication of R dlPFC and mPFC in self-centeredness, which includes not only a personal component (i.e., being the target of someone's greedy action) but also a monetary component (i.e., economic self-interest). In fact, the only condition in the experiment in which we do not observe a decrease in punishment is precisely when a component of self-relevance is absent: the Observer is not the target of the unfairness and he/she also does not bear the costs of punishing the unfair behavior of the Proposer.

Our interpretation is also consistent with findings that lateral prefrontal regions of the brain are also engaged during self-referential tasks, in addition to mPFC and the cortical midline structures often associated with this type of reasoning. This was found to be particularly true in situations with a strong cognitive component, as we believe our task has (Northoff and Bermpohl, 2004; Northoff et al., 2006). Importantly, although the literature on self-referential processing does not investigate economic self-interest per se (i.e., having costs vs. not having costs associated with actions), we argue that bearing monetary costs is an aspect of great relevance to self. Future studies may focus on further unraveling the role of monetary costs on self-referential processing.

The R dlPFC and mPFC on reward

Stimulation of the R dlPFC and anterior mPFC lead to an interesting unanticipated pattern of results, which points to a clear-cut difference between conditions in which rewarding is costly or not. Generally, we observe an increase in non-costly reward following disruption of both the R dlPFC and mPFC. In addition, we observe a reduction of costly reward following inhibition of the R dlPFC mainly when being an unaffected Observer, while inhibition of the mPFC reduces costly reward irrespective of the personal role.

The implication of the R dlPFC in the *cognitive control* mechanism suggests that this brain area is needed to override economic selfish impulses (Knoch et al., 2006). Therefore, we expected that inhibition of R dlPFC would diminish costly reward irrespective of personal involvement. Our results are not consistent with this idea, as we find a decrease in costly reward for Observers but not for Receivers. The lack of differences in costly reward for Receivers following disruption of R dlPFC compared to sham does, however, replicate previous findings showing that inhibition of R dlPFC has no effect on the behavior of a trustee in the Trust Game when interactions are anonymous and building a reputation is not possible (Knoch et al., 2009). Although reputation building is also not possible for Observers in our experiment, we still observe a decrease in costly reward compared to sham. In addition, since we do not find a decrease in reward across all experimental conditions, our results do also not support the *integration-and-selection* mechanism for the dlPFC in the reward domain.

Regarding the role of the mPFC on costly reward, our findings are consistent with the idea that the mPFC is involved in a *cognitive control* mechanism, as stimulation of this brain area leads to a decrease of costly reward independent of personal involvement. Moreover, as stimulation of mPFC affects reward also when being an Observer, we do not find support for the *self-referential* mechanism in the reward domain.

When rewarding is not costly, we observe a strong and unexpected increase in reward after stimulation of the R dlPFC and mPFC for both roles of Receiver and Observer. This finding could be interpreted as an increase in generosity, which in our task would be expressed via rewards. In line with this, a previous study has reported more prosocial behavior after inhibition of the R dlPFC and dorsal mPFC, although prosocial behavior was characterized by giving in a Dictator game (Christov-Moore et al., 2017). However, the generosity interpretation is controversial as less prosocial behavior was observed following cathodal tDCS of the mPFC (Liao et al., 2018).

A more compelling interpretation of our findings in the reward domain is that behavior is driven by two distinct motives in addition to the economic selfish motivation, namely norm enforcement and reciprocity. According to a selfish motive, an individual should never punish or reward others when it is costly to do so. According to the norm enforcement motive, the allocation that is considered the most fair and socially appropriate (the Equal offers in our case) should be the one that is the most rewarded/least punished. Alternatively, according to a reciprocity motive, the more greedy offers are, the more they should be punished, while the more generous offers are, the more they should be rewarded, irrespective of the fairness and social appropriateness consideration. In the punishment domain, both the norm enforcement and the reciprocity motives lead to the same behavioral outcome. That is, according to both motives greedy offers should be punished more than the equal split. However, in the reward domain, it is possible to disentangle norm enforcement and reciprocity motives as the norm enforcement motive predicts that the most rewarded allocation should be the equal split and the reciprocity motive predicts very generous offers should be the most rewarded. This dissociation allows us to test whether TMS affects these two motives differently.

In the absence of costs, we observe that stimulation of the R dlPFC and mPFC promotes a reciprocity mechanism and results in an increased reward of overly generous offers compared to sham stimulation, where participants seem to follow a norm-based behavior and reward the most the allocation which is considered the fairest (i.e., the equal allocation). This pattern of results is consistent with the idea that the stimulation of the R dlPFC and mPFC hinders the enforcement of the norm of an equal split. Thus, the reciprocity mechanism would be the main force driving the decision, resulting in increased rewarding of Equal, Generous and Very Generous offers compared to sham. Reciprocity has up to now been mainly associated with brain regions not targeted in our study such as the insula, anterior cingulate cortex (Cáceda et al., 2017) and right intraparietal sulcus (Bellucci et al., 2017), suggesting that it could have been potentially left unaltered by the brain stimulation. As robustness check of this interpretation, one could expect to find a difference in the effects of brain stimulation between equal vs. generous and very generous offers. Unfortunately, our design was not built to have enough power to conduct such an analysis. Future studies are needed to provide a further test for this explanation and to try to disentangle different norm enforcement motives from reciprocity in the reward domain.

When rewarding is costly, however, stimulation of the R dlPFC and mPFC allows selfish motives to drive the decision, leading to a general decrease of reward. Although this pattern of behavior is present in most of our costly conditions, we note, that it does not happen in the Observer condition following stimulation of the R dlPFC. This finding is not in line with any of the mechanisms considered in this study. Future research should attempt to replicate this finding and clarify whether there are additional components - other than personal involvement and cost - that could be responsible for it.

Fairness and appropriateness evaluations

In each of the three brain stimulation sessions, we observe an inverted u-shape for fairness and social appropriateness ratings, in which equal offers are the ones considered the fairest and the most socially appropriate, while greedy as well as overly generous offers are considered less fair and less socially appropriate. Interestingly, we observe a divergence between fairness and social appropriateness rating in reward behavior, as we observe that the more generous an offer is the more it is rewarded, suggesting that the decision to reward a generous action is not norm based, but reciprocity driven.

We replicate results in the literature showing that fairness ratings are not affected by brain stimulation of the R dlPFC (Knoch et al., 2006; Strang et al., 2015). In addition, we extend this finding by showing that the same holds for the stimulation of the anterior mPFC. This suggests that the mPFC is, just like the R dlPFC involved in implementing reciprocal fairness without influencing fairness judgements[4]. Moreover, we also provide evidence that behavioral changes following brain stimulation are not due to changes in internalized social norms of what is considered socially appropriate behavior.

Sham stimulation

Even in the absence of brain stimulation, our novel within-subjects design can provide insights for behavioral differences in costly and non-costly punishment

[4]We anticipated that fairness evaluations could change if inhibition of mPFC had influenced theory of mind processing. This possibility was pre-registered, although we were agnostic about what would be the direction of the effect on fairness evaluations.

and reward in second and third-party conditions. During sham stimulation, we find that participants punish more when it is free compared to when it is costly, and when they are in the role of the Receiver (i.e., 2nd party conditions) compared to when they are in the role of the Observer. However, the latter effect seems to be driven by the conditions in which punishing is not costly. While some studies have reported that second-party participants tend to punish more strongly than third-parties when punishing is costly (Fehr and Fischbacher, 2004b; Kim et al., 2013; Leibbrandt and López-Pérez, 2012; Pedersen et al., 2013), we find that Observers punish slightly more than Receivers when punishing is costly. We note, however, that the discrepancy between our results and previous literature could be driven by different experimental setups as all these studies employed economic games that differ from ours in at least one dimension[5].

Regarding reward, we find that participants reward less when doing so is costly, while no effect of personal involvement is observed. Previous studies have shown that rewards are frequently used as tools to enforce social norms and cooperation by second-parties (Andreoni et al., 2003; Sefton et al., 2007) and third-parties (Almenberg et al., 2010; Sutter et al., 2009; de Kwaadsteniet et al., 2013). Our study provides evidence that second and unaffected third-parties behave similarly and may ultimately have similar motivations when engaging in reward of generous behavior.

Concluding remarks

To conclude, our study provides several contributions to the research on reciprocal fairness. First, we present evidence of the roles of R dlPFC and mPFC on punishment and reward. Our findings partially support some of the explanations proposed in the literature for the involvement of these regions in reciprocal fairness, while challenging others. Regarding punishment, our results are consistent with a broad perspective of the implication of R dlPFC and mPFC in self-centeredness, which includes both a cognitive control (Knoch et al., 2006) and self-referential processing (Civai et al., 2015; Corradi-Dell'Acqua et al., 2013). However, we note a striking difference in the functions of these prefrontal areas in the reward domain. There, we find that when cost is present, stimulation of the R dlPFC and mPFC allows selfish economic motivations to overtake the desire for norm enforcing or reciprocating, which leads to a general decrease of reward. In the absence of cost, our results are consistent with the idea that reciprocity is the main force driving the decision, resulting in increased rewarding of Equal, Generous and Very Generous offers following brain stimulation compared to sham.

Second, our results also contribute to the debate regarding whether dual-system theories can explain altruistic behavior (Dreber et al., 2016;

[5]Fehr and Fischbacher (2004b) design the third-party condition such that the participant who is still affected by the decision of the proposer cannot punish his proposer, but the proposer of another group. In Leibbrandt and López-Pérez (2012), the observer can punish the proposer and/or the receiver. Pedersen et al. (2013) has also a different design: in the first stage, the subject (receiver or observer) only witness the proposer taking or giving to the receiver. On a second stage, the participant (receivers or observers) becomes the dictator while the previous dictator becomes the recipient. There is no third party in the second stage and the dictator can take or give from the recipient. Finally, in Kim et al. (2013) played the Ultimatum Game as Receivers for themselves or on behalf of a friend or strange. Thus, in the third-party condition the costs of rejecting an offer were not paid by the participants themselves.

Loewenstein and O'Donoghue, 2007; Moore and Loewenstein, 2004; Rand et al., 2014; Schulz et al., 2014). In other words, whether self interest is a spontaneous response, and people need to deploy self-control to behave altruistically; or alternatively, whether prosociality is the intuitive response and self-control is required in order to behave on behalf of one's self interest. In the punishment domain, our results cannot contribute to this debate, as the decrease in costly punishment observed following stimulation of the R dlPFC and mPFC can be interpreted both as an increase in economic selfishness (Knoch et al., 2006) and an increase in generosity (Christov-Moore et al., 2017). On the other hand, the decrease in costly rewards following stimulation of the R dlPFC and mPFC can only be attributed to an increase in economic selfishness in the absence of cognitive control. Hence, in the reward domain our findings are more in line with economic selfishness being an intuitive response.

Third, by being able to establish a direct qualitative comparison between rewards and punishments in reciprocal fairness, we show important evidence of dissimilarities between these two domains. In the absence of brain stimulation, our findings suggest that personal involvement may weigh in differently for punishment and reward, while costs seem to have similar roles in both dimensions, decreasing the frequency of punishment and reward. Furthermore, stimulation of the R dlPFC and mPFC have a different effect on punishment and reward domains. Although reward in second and third-party interactions has received considerably less attention than punishment, a few studies on this participant have already started to show asymmetries between reciprocating a positive versus a negative behavior (Keysar et al., 2008; Offerman, 2002). Future studies will benefit from approaches that enable qualitative and quantitative direct comparisons of behavioral and neural mechanisms in these two domains, revealing the extent of dissimilarities between them.

Lastly, our results indicate that stimulation of the R dlPFC and mPFC lead to a similar pattern of behavior in the punishment domain. The same is observed in the reward domain, although to a lesser extent. A possible explanation for the similar effect following disruption of the R dlPFC and mPFC is that stimulation of one region may have affected the functioning of the whole network involved. We note that although the two investigated regions are sufficiently distant from each other so that TMS effects should be specific and the coil used is known for its good balance between depth and focality (Deng et al., 2013), brain stimulation is not necessarily focal and we cannot rule out network effects. Moreover, the connectivity between the R dlPFC and the posterior mPFC has already been demonstrated to be responsible for normative decisions in second-party interactions (Baumgartner et al., 2011). Although in our study we target the anterior mPFC, evidence of coupling between this region and the right dlPFC has also been reported (Corradi-Dell'Acqua et al., 2013). Future studies should focus on unraveling the functional specificities of these two regions and how they are connected in reciprocal fairness. Reciprocal fairness is a complex behavior that is likely not restricted to a single area of the frontal cortex (Ruff et al., 2013). The combination of fMRI and TMS techniques may be of great value to this endeavor.

5. Appendix

A.1. Payoff Information

In each experimental session, one trial was randomly selected and the three individuals (i.e., Proposer, Receiver and Observer) were paid their earnings in the task in that selected trial. Payoff calculations were made in ECUs and then converted to Euros (1 ECU = €0.10). Proposers' payoff were always calculated in the same way regardless of the condition selected for payment while Receivers' and Observers' payoff were calculated slightly different depending on the condition selected for payment. Table 2.3 summarizes how the payoffs were calculated. In addition to the earnings from the game, TMS participants, who played in the role of Receivers in the Rec_C and Rec_NC conditions and in the role of Observers in the Obs_C and Obs_NC conditions received the participation fee of €7.5 and an additional €5 depending on their answers in the Social Appropriateness task.

Table 2.3: **Payoff calculations**

	Payoffs		
Condition	**Proposers**	**Receiver**	**Observer**
Rec_C	$Ep - A + 5 \times (PR)$	$A - abs(PR)$	Eo
Rec_NC	$Ep - A + 5 \times (PR)$	A	Eo
Obs_C	$Ep - A + 5 \times (PR)$	A	$Eo - abs(PR)$
Obs_NC	$Ep - A + 5 \times (PR)$	A	Eo

Note: Ep stands for the Proposers' endowment of 240 ECUs; A equals the number of ECUs allocated by the Proposer to the Receiver; PR equals the amount of positive or negative ECUs used by the TMS participant to reward or punish the Proposer. Eo stands for the Observers' endowment of 200 ECUs.

A.2. Statistical models for the analysis reported in the paper

Sham stimulation – Punishment and reward

Random-effects Tobit regressions for panel data left-censored at zero and right-censored at +200 with standard errors clustered at the participant level. Only trials in the sham stimulation are considered. The dependent variable is participants' decisions to punish or reward the Proposer when the Proposer's offer was Very Greedy and Greedy (Model 1) or Equal, Generous and Very Generous (Model 2). Dummy predictors coding cost of punishment/reward, personal involvement and the interaction between cost and personal involvement (in table 2.5) are included as within-subject variables. In addition, we controlled for gender, fairness and appropriateness ratings, trial order within an experimental session (i.e., 1-140) and session number (i.e., 1-3). Baseline is composed by Obs_NC condition.

Table 2.4: **Sham main effect**

	(1) Punishment	(2) Reward
Cost	-74.00***	-60.44***
	(3.087)	(3.339)
Receiver	7.920*	-5.114
	(3.104)	(3.314)
Fairness	-7.098**	-22.05***
	(2.433)	(1.323)
Appropriateness	-22.66***	1.872
	(3.758)	(3.000)
Male	11.57	66.67*
	(19.76)	(30.28)
Trail Number	0.0696^{+}	0.0266
	(0.0378)	(0.0407)
Session Number 2	-8.865	-33.80
	(20.89)	(31.80)
Session Number 3	2.384	21.01
	(20.18)	(30.78)
Constant	141.9***	152.1***
	(15.60)	(23.65)
sigma_u		
	40.22***	62.09***
	(6.597)	(9.383)
sigma_e		
	52.77***	65.58***
	(1.240)	(1.453)
N	1329	1983

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.5: **Sham interaction effects**

	(1)	(2)
	Punishment	Reward
Cost	-57.81***	-58.34***
	(4.279)	(4.665)
Receiver	23.02***	-3.178
	(4.193)	(4.478)
Cost x Receiver	-31.88***	-4.233
	(6.034)	(6.589)
Fairness	-6.750**	-22.01***
	(2.403)	(1.324)
Appropriateness	-23.67***	1.776
	(3.721)	(3.004)
Male	11.61	66.70*
	(19.76)	(30.29)
Trail Number	0.0771*	0.0282
	(0.0373)	(0.0408)
Session Number 2	-9.312	-33.79
	(20.90)	(31.81)
Session Number 3	2.301	21.00
	(20.18)	(30.79)
Constant	134.8***	151.1***
	(15.62)	(23.69)
sigma_u		
	40.24***	62.11***
	(6.616)	(9.385)
sigma_e		
	51.98***	65.57***
	(1.222)	(1.453)
N	1329	1983

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

dlPFC and mPFC stimulation – Punishment and reward

Random-effects Tobit regressions for panel data left-censored at zero and right-censored at +200 are run with standard errors clustered at the participant level. The dependent variable is participants' decisions to punish or reward the Proposer when the Proposer's offer was Very Greedy or Greedy (Model 1) or Equal, Generous and Very Generous (Model 2) reported in Table 2.6. Dummy predictors coding stimulation, cost of punishment/reward, personal involvement and the interaction between stimulation and cost, stimulation and personal involvement as well as the three-way interaction between stimulation, cost of punishment/reward and personal involvement are included as within-subject variables. We controlled for gender, fairness and appropriateness ratings, trial order within an experimental session (i.e., 1-140) and session number (i.e., 1-3). Baseline is composed by Obs_NC condition in the sham treatment.

Table 2.6: **Effect of stimulation on Punishment and Reward**

	(1) Punishment	(2) Reward
dlPFC	0.227	22.34***
	(4.739)	(5.262)
mPFC	-1.573	24.89***
	(4.690)	(5.208)
Cost	-59.04***	-59.34***
	(4.783)	(5.358)
dlPFC x Cost	-13.42*	-45.80***
	(6.821)	(7.690)
mPFC x Cost	-15.16*	-40.47***
	(6.795)	(7.596)
Receiver	23.05***	-3.234
	(4.648)	(5.184)
dlPFC x Receiver	-10.19	5.370
	(6.633)	(7.420)
mPFC x Receiver	-7.073	-2.162
	(6.601)	(7.326)
Cost x Receiver	-30.10***	-6.437
	(6.751)	(7.625)
dlPFC x Cost x Receiver	6.991	12.04
	(9.692)	(10.90)
mPFC x Cost x Receiver	10.23	2.448
	(9.632)	(10.83)
Fairness	-4.945***	-14.87***
	(1.215)	(0.725)
Appropriateness	-22.03***	-5.595**
	(1.945)	(1.878)
Male	12.70	77.13**
	(14.52)	(27.59)
Trail Number	-0.0184	0.0815**
	(0.0244)	(0.0278)
Session Number 2	15.67***	0.764
	(2.509)	(2.798)
Session Number 3	15.50***	10.14***
	(2.435)	(2.699)
Constant	124.0***	126.8***
	(8.562)	(15.03)
sigma_u		
	30.62***	58.62***
	(4.468)	(8.405)
sigma_e		
	58.80***	76.28***
	(0.839)	(1.029)
N	3948	5907

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Below the marginal effect on the censored expected value are computed, given that the observed values are censored at 0 and +200. Control variables include fairness, appropriateness ratings, gender, trial number and session number.

Table 2.7: **Marginal effects for altruistic punishment** (model 1 in Table 2.6).

Condition	df	chi2	P>chi2	Contrast	Std.Err.	[95% CI]	
(dlpfc vs sham) Rec_NC	1	4.38	0.036	-8.37	4.0	-16.21	-0.53
(dlpfc vs sham) Rec_C	1	10.23	0.0014	-9.38	2.93	-15.13	-3.63
(dlpfc vs sham) Obs_NC	1	0.00	0.96	0.19	3.96	-7.58	7.96
(dlpfc vs sham) Obs_C	1	6.8	0.009	-8.15	3.13	-14.28	-2.02
(mpfc vs sham) Rec_NC	1	3.41	0.065	-7.26	3.93	-14.97	0.44
(mpfc vs sham) Rec_C	1	7.17	0.007	-7.87	2.94	-13.63	-2.11
(mpfc vs sham) Obs_NC	1	0.11	0.74	-1.31	3.92	-8.99	6.37
(mpfc vs sham) Obs_C	1	10.96	0.0009	-10.18	3.08	-16.21	-4.15
Joint	8	27.25	0.0006				

Table 2.8: **Marginal effects for altruistic reward** (model 2 in Table 2.6).

Condition	df	chi2	P>chi2	Contrast	Std.Err.	[95% CI]	
(dlpfc vs sham) Rec_NC	1	25.89	0.0000	17.67	3.47	10.86	24.48
(dlpfc vs sham) Rec_C	1	1.11	0.29	-2.68	2.53	-7.65	2.29
(dlpfc vs sham) Obs_NC	1	17.46	0.0000	14.27	3.41	7.58	20.97
(dlpfc vs sham) Obs_C	1	15.18	0.0001	-10.46	2.68	-15.72	-5.2
(mpfc vs sham) Rec_NC	1	18.15	0.0000	14.43	3.39	7.79	21.07
(mpfc vs sham) Rec_C	1	6.64	0.0099	-6.55	2.54	-11.53	-1.57
(mpfc vs sham) Obs_NC	1	21.99	0.0000	15.93	3.4	9.27	22.6
(mpfc vs sham) Obs_C	1	7.32	0.007	-7.14	2.64	-12.32	-1.97
Joint	8	70.39	0.0000				

A.3. *Fairness and social appropriateness ratings per condition and per site of stimulation*

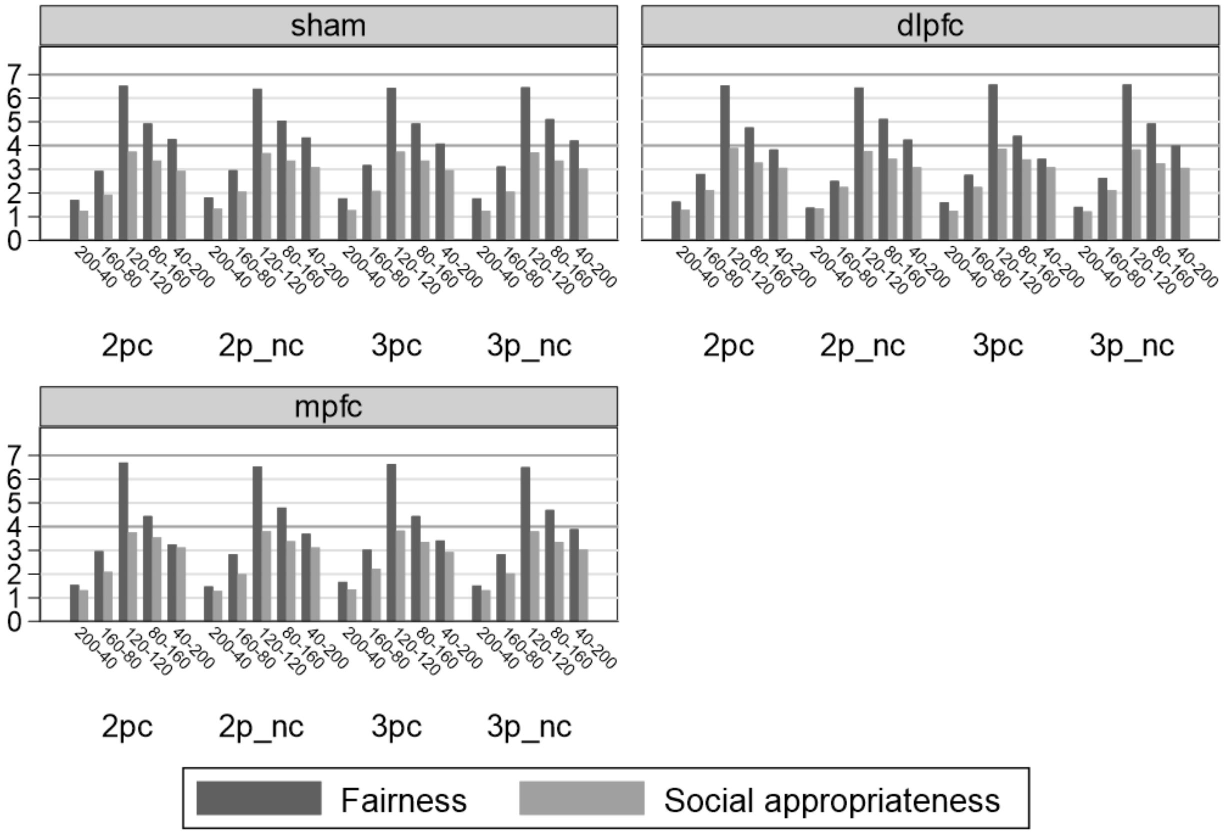


Figure 2.7: **Fairness and social appropriateness ratings** per condition and per site of stimulation. Full sample.

Note: Fairness and social appropriateness ratings are displayed per stimulation and per condition for each allocation bundle available to Proposers. Fairness ratings ranged from 1 (Very unfair) to 7 (Very fair) while Social appropriateness ratings ranged from 1 (Very socially inappropriate) to 4 (Very socially appropriate).

A.4. *Results considering the full sample*

The section below presents the results from the full sample (i.e., weak reciprocators included). Random-effects Tobit regressions for panel data left-censored at zero and right-censored at +200 are run with standard errors clustered at the participant level. Same specifications as reported for the regression conducted with the reduced sample.

Table 2.9: **Effect of stimulation on Punishment and Reward - Full Sample**

	(1) Punishment	(2) Reward
dlPFC	2.482	19.42***
	(4.181)	(4.782)
mPFC	6.345	15.64**
	(4.184)	(4.771)
Cost	-77.55***	-88.37***
	(4.353)	(5.064)
dlPFC x Cost	-7.983	-29.38***
	(6.172)	(7.193)
mPFC x Cost	-16.97**	-24.76***
	(6.206)	(7.192)
Receiver	25.60***	-1.791
	(4.109)	(4.719)
dlPFC x Receiver	-12.26*	-0.798
	(5.845)	(6.749)
mPFC x Receiver	-16.72**	-3.679
	(5.873)	(6.733)
Cost x Receiver	-34.85***	-0.226
	(8.793)	(10.19)
dlPFC x Cost x Receiver	8.889	5.517
	(8.763)	(10.19)
mPFC x Cost x Receiver	22.88**	1.316
	(8.793)	(10.19)
Fairness	-7.330***	-13.90***
	(1.112)	(0.683)
Appropriateness	-20.88***	-13.22***
	(1.694)	(1.652)
Male	12.29	49.81*
	(18.50)	(23.69)
Trail Number	0.0199	0.122***
	(0.0222)	(0.0262)
Session Number 2	14.19***	5.436*
	(2.256)	(2.592)
Session Number 3	10.02***	17.67***
	(2.224)	(2.567)
Constant	121.0***	151.2***
	(10.59)	(13.76)
sigma_u		
	47.04***	60.28***
	(6.145)	(7.527)
sigma_e		
	59.31***	80.64***
	(0.773)	(0.987)
N	5227	7829

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TMS effect on Punishment - Full Sample

Below the marginal effects for punishment (Model 1 in Table 2.9) on the censored expected value are computed, given that the observed values are censored at 0 and +200. Control variables include fairness, appropriateness ratings, gender, trial number and session number.

Table 2.10: **Marginal effects for altruistic punishment.** Full Sample (model 1 in Table 2.9).

Condition	df	chi2	P>chi2	Contrast	Std.Err.	[95% CI]	
(dlpfc vs sham) Rec_NC	1	5.40	0.0201	-7.72	3.32	-14.23	-1.21
(dlpfc vs sham) Rec_C	1	3.52	0.0606	-3.96	2.11	-8.10	-0.18
(dlpfc vs sham) Obs_NC	1	0.35	0.5528	1.94	3.27	-4.46	8.34
(dlpfc vs sham) Obs_C	1	1.40	0.2364	-2.75	2.32	-7.30	1.80
(mpfc vs sham) Rec_NC	1	6.15	0.0132	-8.19	3.30	-14.67	-1.72
(mpfc vs sham) Rec_C	1	0.90	0.3418	-2.04	2.15	-6.24	2.17
(mpfc vs sham) Obs_NC	1	2.30	0.1297	4.97	3.28	-1.46	11.40
(mpfc vs sham) Obs_C	1	5.05	0.0247	-5.18	2.30	-9.70	-0.66
Joint	8	18.31	0.0190				

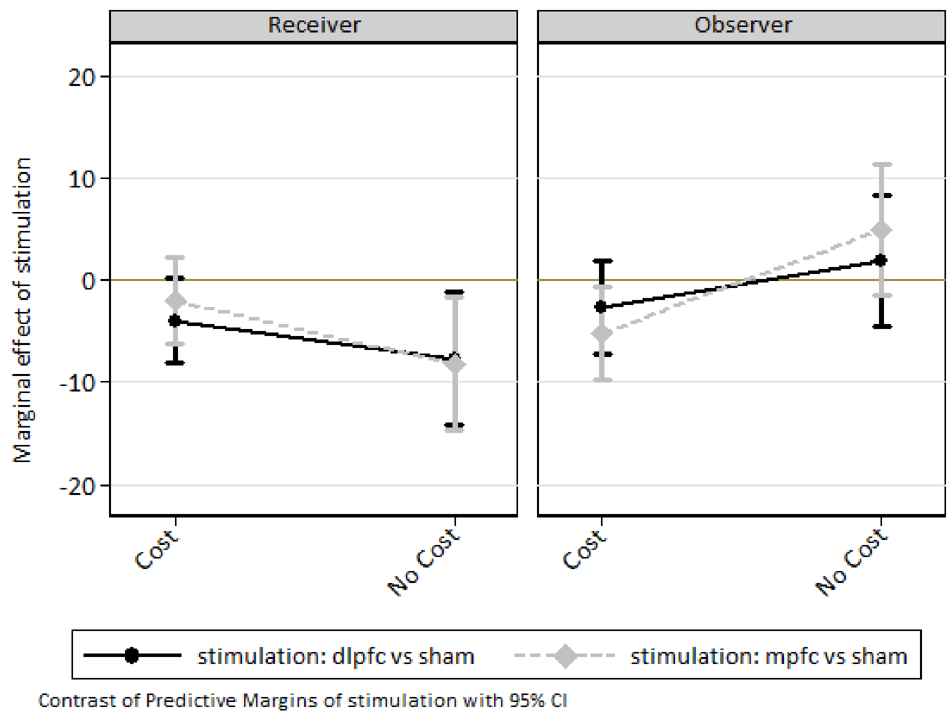


Figure 2.8: **Effect of TMS stimulations on punishment. Full sample.**

Note: The figure summarizes the marginal effect of R dlPFC and mPFC stimulation versus sham stimulation on punishment of Greedy and Very Greedy offers for the whole sample. The horizontal solid line at zero indicates the point estimate for sham treatment in the different conditions. The black circle and the grey diamond markers represent the marginal effects of R dlPFC and mPFC stimulation compared to sham, respectively. The bars indicate the 95% confidence intervals based on the regression analysis.

TMS effect on Reward - Full Sample

Below the marginal effect for reward (Model 2 in Table 2.9) on the censored expected value are computed, given that the observed values are censored at 0 and +200. Control variables include fairness, appropriateness ratings, gender, trial number and session number.

Table 2.11: Marginal effects for altruistic reward. Full sample (model 2 in Table 2.9).

Condition	df	chi2	P>chi2	Contrast	Std.Err.	[95% CI]	
(dlpfc vs sham) Rec_NC	1	14.61	0.0001	11.75	3.07	5.71	17.73
(dlpfc vs sham) Rec_C	1	0.93	0.3353	-2.03	2.10	-6.14	2.09
(dlpfc vs sham) Obs_NC	1	16.20	0.0001	12.27	3.05	6.30	18.25
(dlpfc vs sham) Obs_C	1	3.29	0.0695	-3.83	2.11	-7.97	0.31
(mpfc vs sham) Rec_NC	1	6.13	0.0133	7.49	3.03	1.56	13.42
(mpfc vs sham) Rec_C	1	4.23	0.0398	-4.32	2.10	-8.44	-0.02
(mpfc vs sham) Obs_NC	1	10.62	0.0011	9.86	3.02	3.93	15.79
(mpfc vs sham) Obs_C	1	2.75	0.0975	-3.52	2.13	-7.70	-0.64
Joint	8	40.14	0.0000				

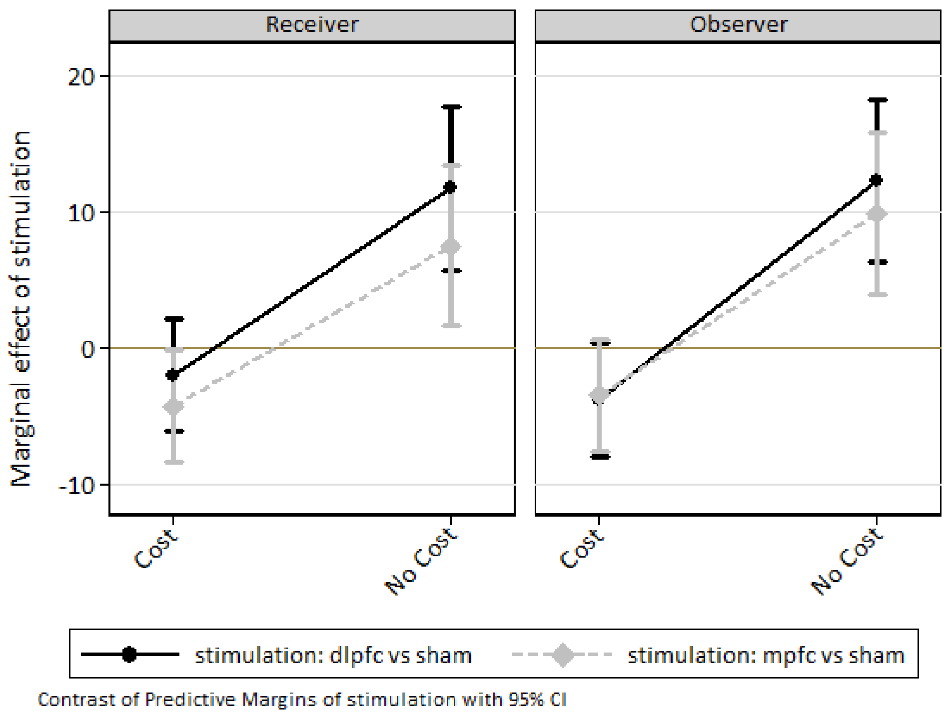


Figure 2.9: **Effect of TMS stimulations on reward. Full sample.**

Note: The figure summarizes the marginal effect of R dlPFC and mPFC stimulation versus sham stimulation on reward of Equal, Generous and Very Generous offers for the whole sample. The horizontal solid line at zero indicates the point estimate for sham treatment in the different conditions. The black circle and the grey diamond markers represent the marginal effects of R dlPFC and mPFC stimulation compared to sham, respectively. The bars indicate the 95% confidence intervals based on the regression analysis.

A.5. *Differences between Weak and Strong reciprocators*

dlPFC Punishment. Differently than the main results reported in the Chapter where we observed a significant decrease in punishment after stimulation of the R dlPFC in the Rec_C condition following Very Greedy and Greedy offers, when considering the full sample this result is only marginally significant ($t = -3.96$, $p = .06$, FDR correction not significant at 90% CI). Furthermore, no differences between right dlPFC and sham stimulation were observed in the Obs_C condition ($t = -2.75$, $p = .23$). Results for the Rec_NC and Obs_NC are the same of the ones reported in the paper.

mPFC Punishment. In the main results reported in the Chapter we observe a significant reduction in punishment of Very Greedy and Greedy offers in the Rec_C condition and a marginally significant decrease in punishment in the Rec_NC condition following downregulation of the mPFC. When considering the full sample we observe no differences between mPFC and sham stimulation in the Rec_C condition ($t = -2.04$, $p = .34$). Moreover, we find a significant decrease in punishment in the Rec_NC condition compared to sham stimulation ($t = -8.19$, $p = .01$, FDR corrected at 95% CI). No differences in results between the full and partial sample were found for the Obs_NC and Obs_C conditions.

dlPFC Reward. Regarding reward, while I report in the Chapter a significant decrease in reward of Equal, Generous and Very Generous offers following stimulation of the R dlPFC compared to sham stimulation in the Obs_C condition, this result is only marginally significant when considering the full sample of participants ($t = -3.83$, $p = .07$, FDR corrected at 90% CI).

mPFC Reward. Differences between the full and partial sample were only found for the Obs_C condition. While we observe a significant decrease in reward of Equal, Generous and Very Generous offers in the Obs_C condition following mPFC stimulation when considering the partial sample, this result in only marginally significant for the full sample of participants ($t = -3.52$, $p = .097$, FDR corrected at 90% CI).

A.6. *Frequency of Reciprocation by Weak and Strong Reciprocators*

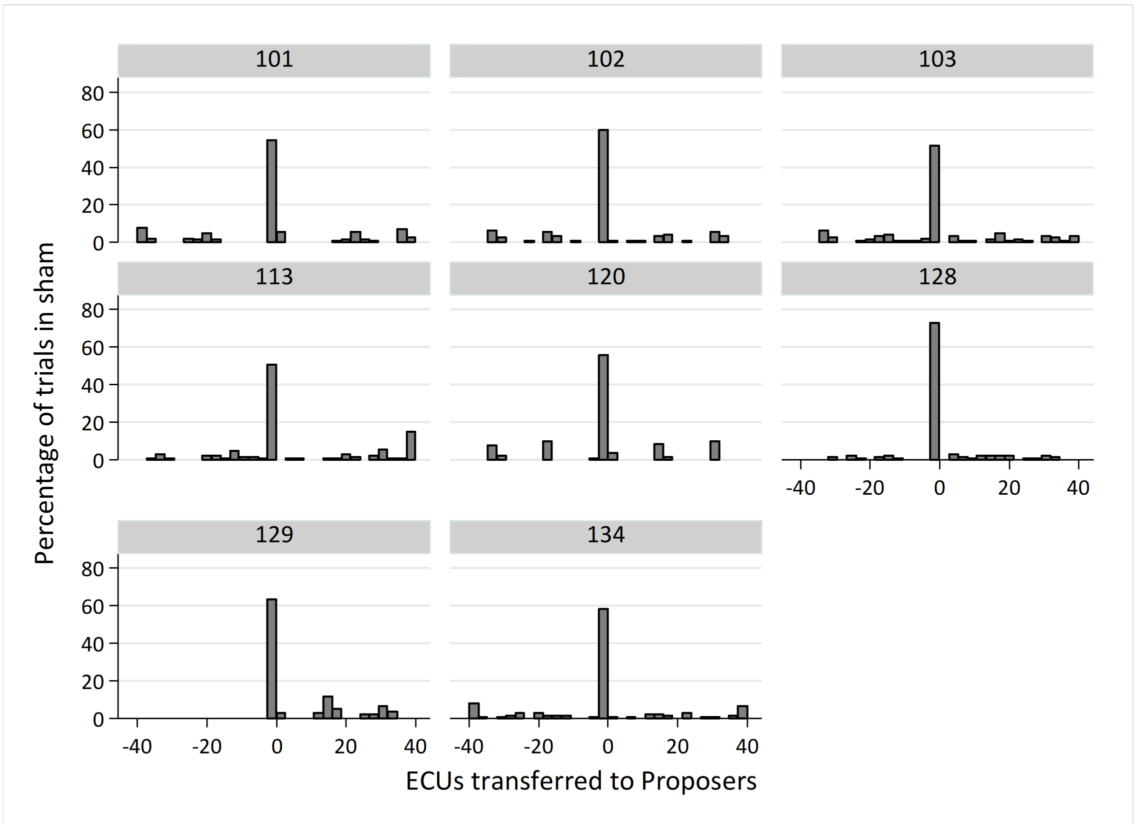


Figure 2.10: **Relative frequency of reciprocity** (punishment and reward) in trials under sham stimulation for weak reciprocators.

Note: The figure displays the frequency to which weak reciprocators chose to transfer negative ECUs to Proposers (punishment) and positive ECUs to Proposers (reward) across all four experimental conditions under sham. Participant IDs are displayed at the top of each panel, Y axis represents the percentage of trials in sham stimulation (collapsed across the four experimental conditions) and the x axis represents the number of ECUs transferred to Proposers (negative ECUs represent punishment, while positive ECUs represent rewards). For these participants, frequency of non-reciprocity was high, as they chose to not punish or reward the Proposer in more than 50% of trials.

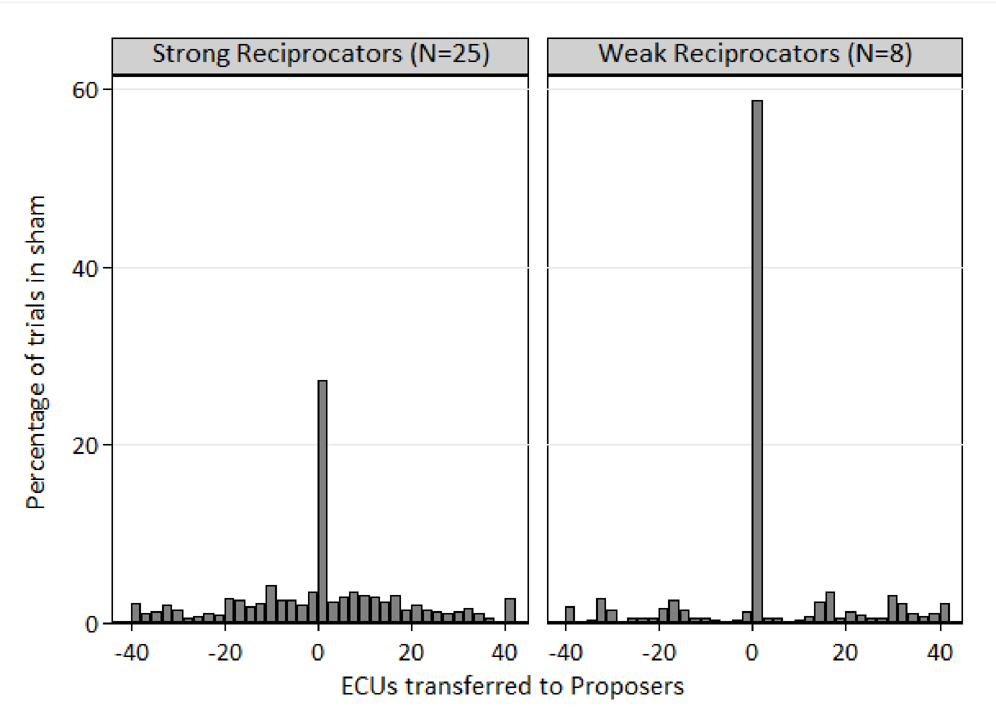


Figure 2.11: **Relative frequency of reciprocity** (punishment and reward) in trials under sham stimulation for strong and weak reciprocators.

Note: The figure displays how weak and strong reciprocators chose to respond to Proposers. The left panel displays the pattern of response for 25 participants who punished or rewarded Proposers in at least 50% of the trials (strong reciprocators), while the right panel displays the pattern of responses of all participants who chose to not punish or reward Proposers in more than 50% of trials (weak reciprocators). The y axis represents the percentage of trials in sham stimulation (collapsed across the four experimental conditions) and the x axis represents the number of ECUs transferred to Proposers (negative ECUs represent punishment, while positive ECUs represent rewards).

A.7. Effect of TMS for weak reciprocators.

For completeness, this section presents the results when considering only the Weak Reciprocators who are excluded from the main analysis presented in the main text of the paper because they did not punish or reward more than 50% of the times following sham stimulation. Importantly, we report below the effect of brain stimulation for this subset of participant to check whether TMS over dlPFC and mPFC changes their behavior compared to sham. Results show that TMS over dlPFC and mPFC does not change punishment of Greedy and Very Greedy offers. It also does not change reward of Equal, Generous and Very Generous offers. This indicates that, as expected, brain stimulation did not affect the behavior of Weak Reciprocators. Table 2.12 shows the result of a random-effects Tobit regressions for panel data left-censored at zero and right-censored at +200 with standard errors clustered at the participant level. The dependent variable is participants' decisions to punish or reward the Proposer when the Proposer's offer was Very Greedy or Greedy (Model 1) or Equal, Generous and Very Generous (Model 2). Baseline is composed by the sham treatment.

Table 2.12: **Effect of TMS for Weak reciprocators only.**

	(1) Punishment	(2) Reward
dlPFC	-7.029	10.92
	(8.415)	(9.076)
mPFC	4.409	8.550
	(8.904)	(9.585)
Fairness	-36.87***	22.43***
	(2.817)	(2.508)
Appropriateness	-25.01***	-5.625
	(4.677)	(4.960)
Constant	63.42^{+}	-183.3***
	(35.89)	(19.20)
sigma_u		
	97.22***	34.66***
	(27.12)	(9.565)
sigma_e		
	124.0***	162.0***
	(4.178)	(4.986)
N	2949	3034

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6. Instructions

Welcome to this decision-making experiment. [6] The webcam on top of your computer screen will be turned on and you will be filmed for part of this experiment. You will be informed when the camera will be turned on and when it will be turned off. The video will be seen only by the experimenters and will be destroyed once data have been gathered from it. The content of the video will not affect your payment. If you disagree with being filmed, you are free to leave at any time. In this case, you will receive the show-up fee, but you will not receive additional payment.

While being filmed, we kindly ask you to pay attention to the following aspects:

- Please do not touch the camera or move the screen. If you think there is a problem with the camera, please inform the experimenter.
- Please avoid placing your hands on your face.
- Please remove any caps or hats.
- Remain seated properly and please avoid leaning on the chair.
- Please do not turn away from your screen.

For your participation in this experiment, you will receive a fixed participation fee of €7.5. In addition, you can earn money with the decisions you make. Please read these instructions carefully. Importantly, we employ a **strict non-deception policy**. This means that all information you receive from this experiment is truthful and when you are told to interact with other participants these other participants do actually exist.

This experiment consists of a Practice Part, followed by a Task Part. The Practice Part consists of these instructions and some comprehension questions. It is important that you answer the comprehension questions correctly. When you have answered all comprehension questions correctly you can proceed to the Task Part.

In the Task Part, you **can earn money** by making choices. These earnings depend on your decisions and on the decisions of other participants of this experiment. This will be explained in detail later in these instructions. In the experiment, we will use **ECU (Experimental Currency Units)**, which will be converted to Euro at the end of the experiment. For all the participants in this experiment, each ECU will be converted to Euros at a rate of 1 ECU to €0.10.

If you have any questions during the experiment, please raise your hand and an experimenter will come to you to answer your question in private.

Description of the decision situations

In this experiment, you have to make choices in **different decision situations**. Each decision situation will be different and it will be explained in detail. In each decision situation, there are three different participants, randomly matched with each other (Figure 2.12). The pairing is anonymous, meaning that neither individual will ever know the identity of the other individuals with whom he or she is matched. One participant will be known as **Person A**, one participant will be known as **Person B** and the other participant as **Person C**.

In each decision situation, Person A allocates money between him/herself and Person B. Each decision situation consists of two stages:

The first stage

At the beginning of the first stage:

- Person A has 240 ECU at his/her disposal.
- Person B has 0 ECU at his/her own disposal.

[6] The text font, the size and the appearance of images have been adapted from the original instruction version.

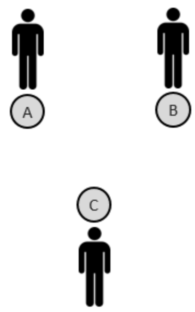


Figure 2.12: Participants in each decision situation.

- Person C has 200 ECU at his/her disposal

Person A decides how to distribute his/her 240 ECU between him-/herself and Person B. Person A is always represented by a red stylized human (Figure 2.13). When making the decision, Person A can choose one of the following options:

- Keep 200 ECU ($\pm$ 3 ECU) and Give 40 ECU ($\pm$3 ECU) to Person B
- Keep 160 ECU ($\pm$3 ECU) and Give 80 ECU ($\pm$3 ECU) to Person B
- Keep 120 ECU ($\pm$3 ECU) and Give 120 ECU ($\pm$3 ECU) to Person B
- Keep 80 ECU ($\pm$3 ECU) and Give 160 ECU ($\pm$3 ECU) to Person B
- Keep 40 ECU ($\pm$3 ECU) and Give 200 ECU ($\pm$3 ECU) to Person B

The indicated $\pm$ 3 ECU means that, for instance, if Person A chooses the fourth option then it may be that Person A kept anything between 83 and 77 ECU and that Person A gave anything between 163 and 157 ECU. The exact amount is randomly decided. Equivalently for the other options.

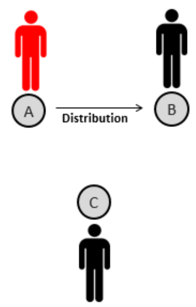


Figure 2.13: Person A's action.

The second stage

Your Role. In some decision situations you will be Person B and in some decision situations you will be Person C. You will never be Person A. You will always be represented by a blue stylized man in a rectangular shape.

In each decision situation, you will learn about Person A's choice. Thereafter, you can make the decision to increase, decrease or leave unchanged Person A's payoff. How exactly this can be done will be described in details later in the instructions.

In each decision situation, before you make your decisions, you will be informed whether you are Person B or Person C.

- If you are Person B: Figure 2.14 depicts a decision situation where you are Person B. In this type of situation, Person A has chosen how to allocate the 240 ECU at his/her own disposal between him/herself and you. After having learned about Person A's allocation, you can decide to increase, decrease or leave unchanged Person A's payoff. In this case, Person C, who is another participant of this experiment, has 200 ECU and will only be informed about Person A's decision, but he/she will not know what you decide. Person C cannot modify yours or Person A's payoffs.

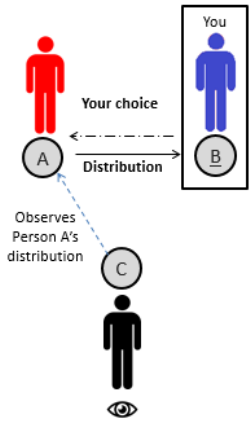


Figure 2.14: Decision situation in which you are Person B.

- If you are Person C: Figure 2.15 below depicts a decision situation where you are Person C. In this case, Person A has chosen how to allocate the 240 ECU at his/her own disposal between him/herself and Person B, who is another participant of this experiment. After having learned about Person A's allocation, you can decide to increase, decrease or leave unchanged Person A's payoff. In this case, you receive 200 ECU at the beginning of the decision situation. Person B will only be informed about Person A's decision, but he/she will not know what you decide. Person B cannot modify yours or Person A's payoffs.

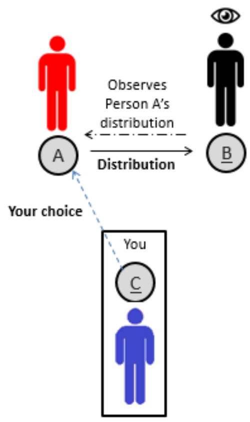


Figure 2.15: Decision situation in which you are Person C.

Your decision

In some decision situations your decisions can be costly and in some other decision situations, they will not be costly. In each decision situation, before having to make a choice, you will be informed whether your decision is costly or non-costly.

- Costly decisions: In some decision situations, you will have to use some of your ECU to modify Person A's payoff (Figure 2.16). Specifically:
  - if you choose to increase Person A's payoff, you pay 1 ECU for each 5 ECU by which Person A's payoff is increased,
  - if you choose to decrease Person A's payoff, you pay 1 ECU for each 5 ECU by which Person A's payoff is decreased,
  - if you choose to leave unchanged Person A's payoff, you do not pay any ECU.

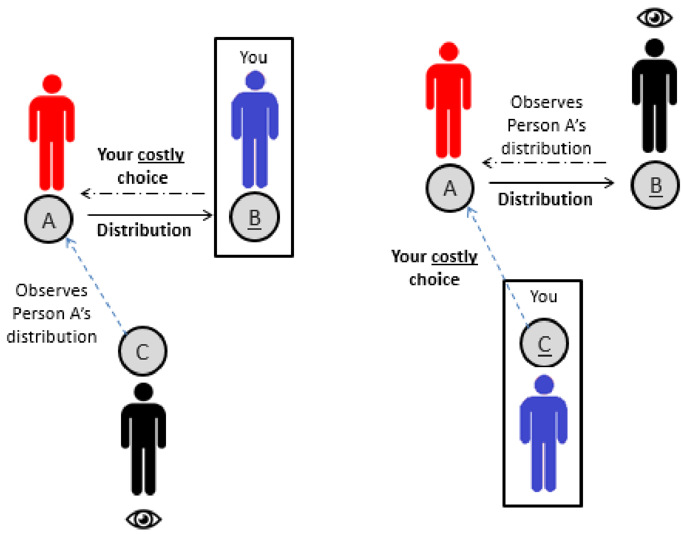


Figure 2.16: Costly decisions. In this type of decision situations, you pay with your ECU to modify Person A's payoff.

- Non-Costly decisions: In some decision situations, you will not have to use your ECU to modify Person A's payoff (Figure 2.17). Specifically:
  - if you choose to increase Person A's payoff, you pay 0 ECU for each 5 ECU by which Person A's payoff is increased,
  - if you choose to decrease Person A's payoff, you pay 0 ECU for each 5 ECU by which Person A's payoff is decreased,
  - if you choose to leave Person A's payoff unchanged, you do not pay any ECU.

Information

In a decision situation, each Person is informed regarding the options of the other participants he/she is matched with. The rules are therefore common knowledge. For instance, before choosing how to allocate the 240 ECU, Person A knows which are the possible actions available to Person B and Person C and which information they receive.

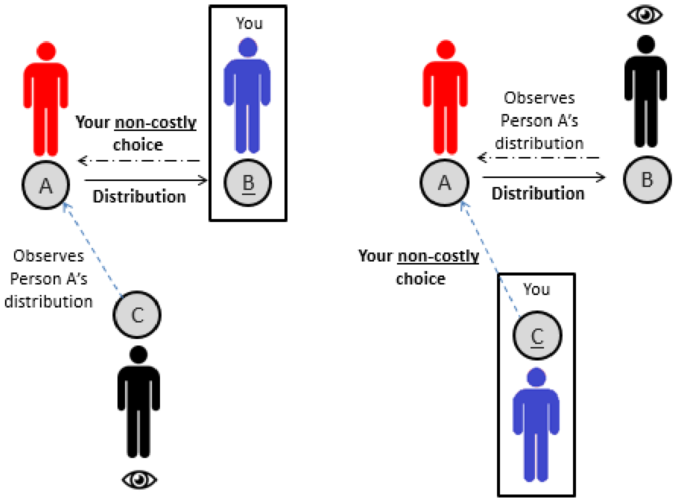


Figure 2.17: Non-costly decisions. In these type of decision situations, you will not pay any ECU to modify Person A's payoff.

Condition announcement

During the experiment, you first receive information regarding your role (i.e., Person B or Person C) and the cost of your choice (i.e., Cost or no Cost) as shown below (Figure 2.18). Thereafter, you will be presented with several decision situations where your role and the cost of your choice remain constant. You will be informed when your role and/or the cost of your decision change again.

For instance, in the example in Figure 2.18 you are informed that in the next multiple decision situations you will be Person C and that you pay 1 ECU for each 5 ECU by which Person A's payoff is increased or decreased (i.e., Costly decision). Be aware that you will face many decision situations in this experiment and that there will be waiting periods.

Screen example

After seeing the condition announcement, you will face the decision situations. You can see below an example of the screen of a decision situation you will face during the experiment (Figure 2.19). Please note that for each decision situation, you will be matched with two new participants. This means you will never be exposed to the same participants more than once.

- On the top right part of the screen, you can see the remaining time you have to make your decision. You have at most 20 seconds to make your decision and click on the "Confirm" button. If you do not make a decision within this time limit you will automatically move to the next decision situation.
- In the centre of the screen, the image with the stylized humans summarizes the features of the decision situation you are facing. This picture is a reminder of the condition announcement you saw before starting the decision situations. You are informed about your role represented by a blue human (i.e., Person B or Person C) and about the cost of your decision (i.e., cost or no-cost). You will be presented with several decision situations where your role and the cost of your decision remain constant. You will be informed when your role and/or the cost of your decision vary.
- Above the picture, a two-letter ID associated to the Person A you are matched

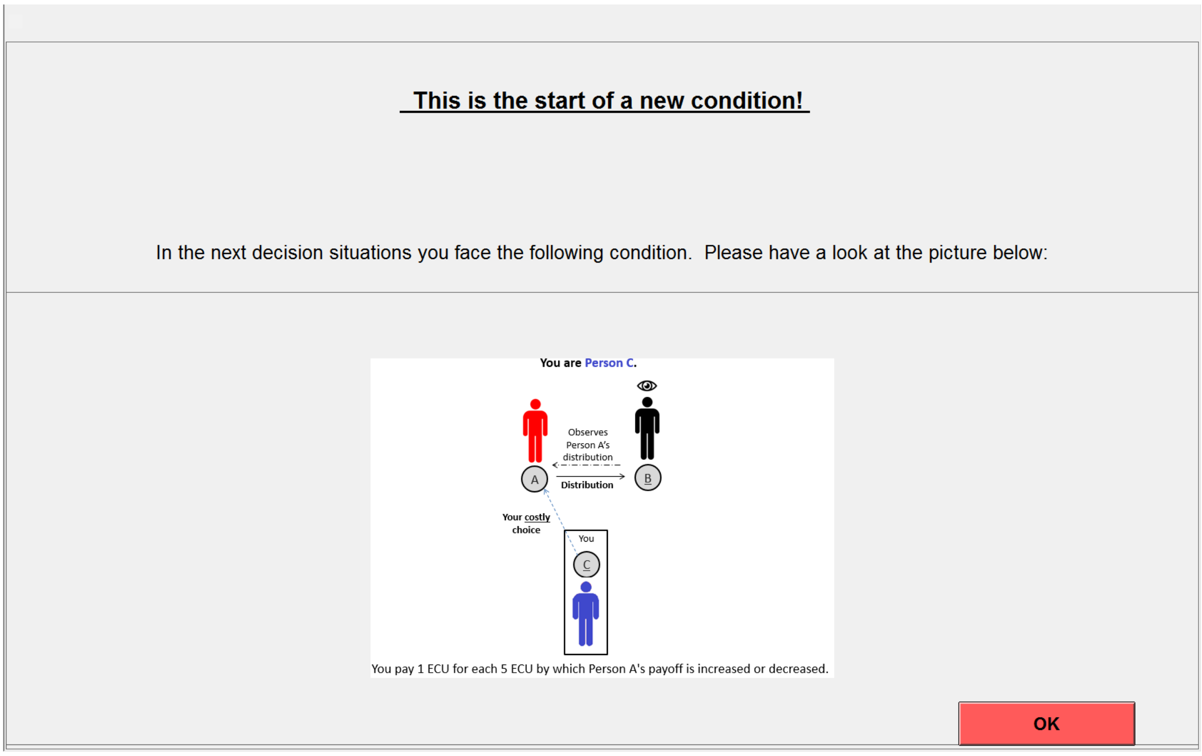


Figure 2.18: Condition announcement. You are informed about your role and the cost of your choice for the following decision situations.

with in that decision situation is displayed. This ID changes in each decision situation, since you are always matched with a different Person A.

- Below the picture, you learn about Person A' decision to distribute his/her tokens with Person B. Please pay particular attention to this because it might vary across decision situations.
- Your decision consists of moving a slider to decrease or increase Person A' payoff. If you move the cursor to the left, you will decrease Person A's payoff. If you move the cursor to the right, you will increase Person A' payoff. You can move the cursor in the slider or click on the button on the left and on the right. You can use up to 40 ECU of your own to modify Person A's payoff. If you do not move the cursor or click on any button, you leave Person A's payoff unchanged.
- On the bottom of the screen, you are informed about your payoff and Person A's payoff. When you move the slider, this payoffs will automatically update.
- You need to click on Confirm to submit your decision.

Your identity

Your identity and the identity of all other participants is secret. You will never be asked to reveal it to the other participants during the course of the experiment or after it. Your name is recorded only at the end of the experiment to pay you. The other participants will not be able to link your identity to any of the decisions you make. In order to keep your decisions private, please do not reveal your choices to any other participant. All the interactions are anonymous and none of the participants can communicate or influence the choices of the others.

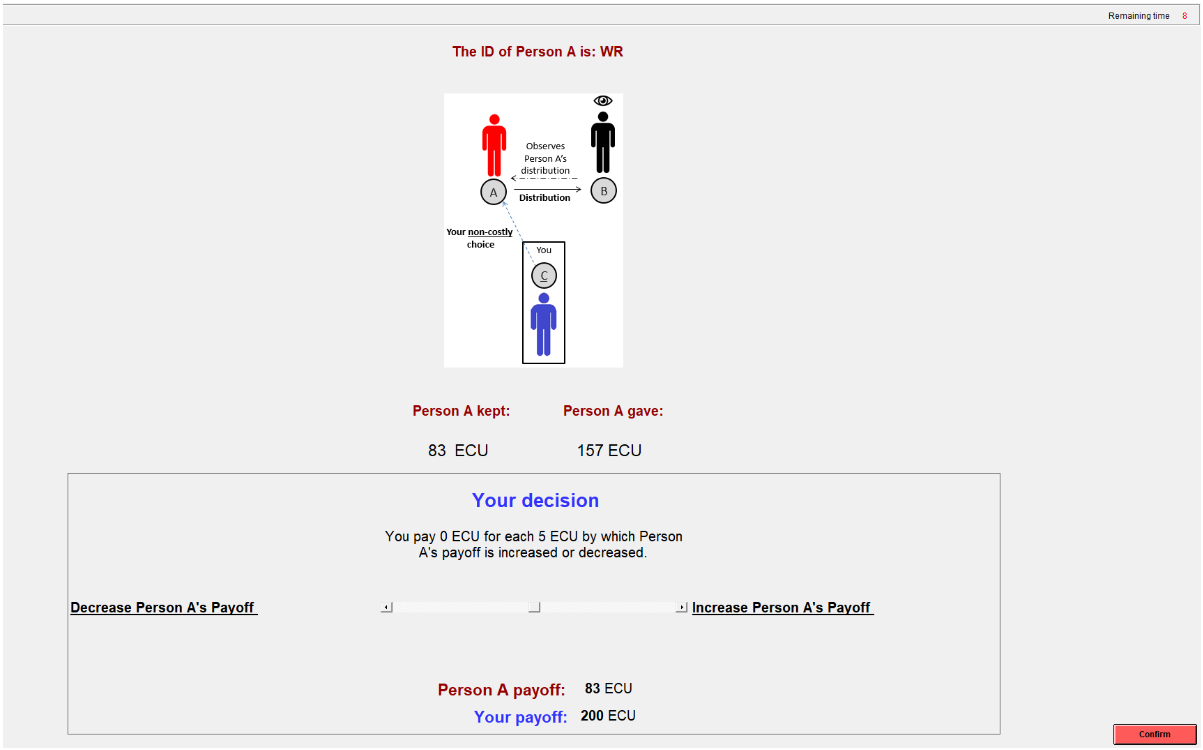


Figure 2.19: Screen Interface example.

Matching

Importantly, in this experiment, you will never face the same participant more than once. Thus, you should consider each decision you make as completely independent from the other decisions: in each decision situation, you will be matched with different other participants. The actions of these participants have been recorded in a previous experimental session.

Payment

For each experimental session, you receive a show up fee of €7.5. In addition, in each experimental session, one of the decision situations you faced will be randomly selected for payout. You and the other two Persons who participated in the selected decision situation will be paid according to Person A's and your choices. Please note that if you did not make a decision within the time limit for the selected decision situation that counts for payout, you receive €0 while the other two persons will be paid according to the initial ECU allocation. All your earnings, including the €7.5 participation fees, will be transferred to your bank account at the end of the three experimental sessions. Your earnings will be transferred within one week from the end of the experiment.

Note: Recall, for each participant in this experiment, each ECU will be converted to money at a rate of 1 ECU to €0.10.

This is the end of the instructions. You can now answer the comprehension questions. If you have any questions please raise your hand and wait for the experimenter.

Chapter 3

Intertemporal Social Preferences

Abstract

People are often confronted with situations that involve a trade-off between immediate and delayed benefits and costs, and between one's own and others' welfare. Despite the importance of these decisions, intertemporal and social domains are traditionally analyzed separately. In this paper, we use modified Dictator Games and examine the relationship between time delay and generosity in a laboratory experiment. In a between-subjects design, we vary the time of payout for the dictator and the receiver. Within-subjects, we vary the endowment of the dictator as well as the price of giving and keeping and the resulting relative price of giving. Our results show that, at the aggregate level, time delay does not affect giving. This result holds also when taking into account the differences in discount rates for own and the recipient's payoffs. However, we observe that time delay affects the sensitivity to different prices of giving. Only when both dictator and recipient receive their payoff immediately, is giving sensitive to price changes. Third, our data suggest that the finding of Andreoni and Vesterlund (2001), that men are more sensitive to the price of giving than women, extends to contexts with delayed payments. This also holds for the finding that women give more than men when giving is expensive, but not for the finding that men are more generous than women when giving is cheap. Finally, we observe that giving decisions largely respect GARP also in the presence of delayed payments, suggesting that choices can be rationalized by a well-behaved utility function.

This paper is co-authored with Arno Riedl, Giang Tran & Matthias Wibral.

1. Introduction

A large body of evidence shows that many people have social preferences, i.e., they care not only about their own payoff, but also the payoff of the persons they are interacting with (e.g., Charness and Rabin, 2002; Chen and Li, 2009; Fehr and Fischbacher, 2002; Camerer and Fehr, 2004). Such preferences have important consequences for behavior in a wide range of contexts and domains ranging from incentive systems or negotiations to charitable giving. So far, the literature on prosocial behavior has typically studied allocation decisions between an individual and another person, in which the decision and the consequences occur at the same time. However, in many contexts, if not most, prosocial behavior has an intertemporal dimension. For instance, the costs of donating to a charitable organization are often immediate, but the benefits to the ultimate recipient are delayed.[1] Donors might also pledge a donation in the future, postponing both the cost and the benefit (Andreoni and Serra-Garcia, 2020). Despite the prevalence and importance of this type of decisions, intertemporal and social preferences are traditionally analyzed separately and surprisingly little research has been devoted to understanding intertemporal social decision making (for exceptions see Kovarik, 2009; Dreber et al., 2016; Kölle and Wenner, 2018; Rong et al., 2019). The aim of this paper is to provide an in-depth analysis of such decisions using a laboratory experiment.

To put structure on our analysis while keeping it tractable we consider the standard workhorse for studying prosocial behavior, the Dictator Game, in four different treatments. The treatments vary at which of two points in time (now or in five weeks) the payment to the dictator or the receiver, respectively, is made. We analyze these differences in a between-subjects design. For a more complete picture of intertemporal social preferences, we additionally vary within-subjects the dictator's endowments, the price of giving, and the price of keeping (in the spirit of Andreoni & Miller, 2002). In addition, we elicit both the discount rate for own and the recipient's payoffs separately using the convex budget method (Andreoni and Sprenger, 2012). This allows to test whether dictators discount the utility received from own payoff and the utility received from the other's payoff at different rates and to analyze how time preferences relate to giving in the different treatments.[2]

This approach has several advantages. First, it allows us to study whether the sensitivity of giving to different prices changes with time delay. This is especially important in the context of matching grants where a company or institution promises to match the donations from individual donors. For example, a 1:1 matching effectively halves the relative price of giving. In our setup, we can study whether such changes in prices become less or more effective when the money reaches the recipient with a delay. Second, we are able to check whether the important finding by Andreoni and Vesterlund (2001) that men and women have different demand curves for altruism replicates, and extends to settings with an intertemporal dimension. Third, we

[1] On a larger time scale one could think of carbon reduction efforts in this context. Adhering to a carbon reduction agreement is costly in the short run for the current generation, while its benefits will materialize in the future and mostly benefit other generations.

[2] For instance, if the dictator discount the utility received from her payoff more than the utility received from the other's payoff, one could predict that she will give more of her endowment when both she and the recipient receive their payoff later compared to a situation where both receive their payoff earlier.

can test whether time delay affects the rationality of giving decisions according to the Generalized Axiom of Revealed Preference (GARP). Given the additional layer of complexity, we might expect a lower degree of rationalizability in treatments which involve time delay.

Our main findings are as follows. First, we do not find any significant differences in giving between treatments when aggregating over all endowments and prices. Dictators give around 15% of their endowment irrespective of the timing of payments. Second, we do observe, however, that the timing of payments affects the sensitivity to different prices. Only when both dictator and recipient receive their payoff immediately, does giving increase with the relative price of giving. This relationship holds also when looking at the prices of keeping and the price of giving separately. That is, dictators give more (less) when keeping (giving) is costly. In the other treatments that involve time delays, dictators are not sensitive to changes in the prices. Third, our data suggest that the finding of Andreoni and Vesterlund (2001) that men are more sensitive to the price of giving than women extends to contexts with delayed payments. This holds also for the finding that women give more than men when giving is expensive, but not for the finding that men are more generous than women when giving is cheap. Fourth, giving decisions respect GARP also in the presence of delayed payments, indicating that choices can be rationalized by a well-behaved utility function.

Our study differs from previous work in several important dimensions. First, we focus on a comprehensive set of combinations of immediate and delayed payments instead of comparing only a subset of these. Second, by explicitly measuring discount rates for self and other we are able to test potential channels through which time delay might affect giving behavior. Third, we study behavior for a wide range of prices of keeping and giving. We are thus able to investigate differences in rationalizability of choices. We extend the analysis of the sensitivity of giving to different prices to contexts involving time delays. Finally, our findings contribute to the literature on discounting for others.

2. Literature review

Our study is related to several strands of the literature. Most importantly, it contributes to the nascent literature on intertemporal social preferences by providing a comprehensive analysis and empirical test. A few paper compare giving in a Dictator Game when payments for *both* the dictator and the recipient occur at a later point in time compared to a situation in which both are paid immediately. The evidence so far is mixed. Both Kovarik (2009) and Dreber et al. (2016) find that giving is lower when payments are postponed for both the dictator and the recipient. The opposite finding, i.e., higher giving when both payments are delayed is reported in a hypothetical Dictator Game study by Yi et al. (2011).[3] Higher giving when both payments are delayed, has also been found in two studies on giving to charitable organizations. In a natural field experiment, Breman (2011) investigates how donors to a charitable organization react when asked to increase their monthly contributions, either immediately or at a later date. The findings showed

[3]In line with this, Agerström and Björklund (2009) find that people report that they would be less likely to choose selfish over altruistic behaviors when thinking about distant compared to near future events.

significantly higher donations when dictators are asked to commit to future donations. Andreoni and Serra-Garcia (2020) examine how donations are affected when they are delayed and individuals commit in advance to their giving decisions. More individuals decide to give when the decision is made in advance, than when both the decision and payment take place immediately. Andreoni and Serra-Garcia (2020) suggest that individuals who dislike giving, choose not to give when the gift is paid immediately, but might choose to give when the cost of giving is delayed.

In a hypothetical study, Rogers and Bazerman (2008) show that people are more likely to forgo an amount of money in order to donate it to a charity when their choices are implemented in the distant future rather than immediately. Another hypothetical study (Ida and Ogawa, 2012) finds lower giving when the dictator receives the payoff immediately whereas the receiver's payoff is delayed, compared to a situation in which both receive their payoff immediately. Our paper is also closely related to two recent papers by Kölle and Wenner (2018) and Rong et al. (2019). Kölle and Wenner (2018) study whether individuals exhibit present bias in a task where they have to allocate real effort between themselves and others at different points in time. In contrast to most of the literature, they assume that individuals discount payoffs, and not utility. Using this model, they find substantial present bias in generosity and a correlation between present bias in social and non-social contexts at the individual level. Rong et al. (2019) study a framework in which the discount rate a person uses in intertemporal decisions depends on both the recipient at the earlier point in time (herself or someone else) and at the later point in time (herself or someone else). This implies that in addition to the conventional discount rates for self and other, they assume that individuals have two additional specific discount rates: one for situations involving an earlier payoff for the self and a later payoff for the other, and one for situations involving an earlier payoff for the other and a later payoff for the self. Using the convex time budget method they estimate the resulting four different discount rates. Their main finding is that their participants have a higher discount rate in situations in which the earlier payment is for themselves and the payment at the later point in time for someone else compared to situations in which both payments are for themselves or both payments are for someone else.

3. Methods

In this section, we describe the different parts of the experimental design. First, we explain in detail the modified version of the Dictator Game we implemented, including the different prices of giving and prices of keeping. Second, we present the different treatments where we vary the time delays associated with the payoffs for the dictators and/or the recipients. Thereafter, we describe the task we use to elicit intertemporal preferences of the dictators for themselves and for others. Finally, we describe the experiment procedures.

3.1. Modified Dictator Game

Dictators played a series of Dictator Games (Andreoni and Miller, 2002) in a randomized order that varied the endowment that could be redistributed, the prices of giving (p_{give}) and keeping (p_{keep}) and, as a consequence, the relative price of giving k which is defined as $k = p_{give}/p_{keep}$. Table 3.1 shows the 15 decision situations each dictator faced during the experiment. Decision situations differ in the number of tokens to be divided with an anonymous recipient and in the number of points a token was worth to each player. The total amount of tokens to be divided is 40, 60, 72, 90 or 120 and tokens have a hold or pass value of 1, 2, or 3 points each. This difference in values generates situations where giving and keeping are more or less costly. As in Andreoni and Vesterlund (2001) and Andreoni and Miller (2002), we define the inverse of the Hold Value is the price of self-payoff p_{keep}, and the inverse of the Pass Value is the price of other payoff p_{give}.

For instance, in Situation 1 the prices are $p_{keep} = 1/3$ and $p_{give} = 1$. Hence, transferring one token increases the recipient's payoff by one point, but decreases the dictator's payoff by three points. The situations where $p_{keep} = p_{give} = 1$ are standard Dictator Games. In situations where $p_{give} < p_{keep}$, passing one token decreases the dictator's payoff by one point, but increases the recipient's payoff by either two or three points. At the end of the experiment, each point earned is converted to €0.10 and paid to the dictators and recipients, respectively.

Table 3.1: **Dictator Game. Experimental parameters**

Situation	**Endowment**	**Hold value**	**Pass value**	**p_{keep}**	**p_{give}**	**Relative price *k***
1	40	3	1	1/3	1	3
2	40	1	3	1	1/3	1/3
3	60	2	1	0.5	1	2
4	60	1	2	1	0.5	0.5
5	60	3	1	1/3	1	3
6	60	1	3	1	1/3	1/3
7	60	1	1	1	1	1
8	72	2	1	0.5	1	2
9	72	1	2	1	0.5	0.5
10	72	3	1	1/3	1	3
11	72	1	3	1	1/3	1/3
12	72	1	1	1	1	1
13	90	1	2	1	0.5	0.5
14	90	2	1	0.5	1	2
15	120	1	1	1	1	1

Note: The table describes the 15 decision situations which vary the resources that could be redistributed (i..e, Endowment), p_{keep}, p_{give} and the relative price of giving *k*.

3.2. Time delay treatments

We implemented a between-subjects design where the participants are allocated to one of four time-delay treatments. In each treatment, a dictator decides how to share the endowment with a recipient after being informed about the timing when she and the recipient will receive the payoffs. The four treatments are listed below, where the first term refers to the delay associated with the payoff of the dictator and the second term to the delay associated with the payoff of the recipient, respectively:

- *NowNow*. Both dictator and recipient receive their earnings on the day of the experiment.
- *NowLater*. The dictator receives the earnings on the day of the experiment, while the recipient receives the earnings 35 days after the experiment took place.
- *LaterLater*. Both dictator and recipient receive their earnings 35 days after the experiment took place.
- *LaterNow*. The dictator receives the earnings 35 days after the experiment took place, while the recipient receives them on the day of the experiment.

Table 3.2 shows the different time delays for the payoff of dictators and recipients, in each of the four treatments. The last column shows the number of dictators participating in each treatment.[4]

Table 3.2: **Dictator Game. Time delays.**

Treatment	**Payoff dictator** *(days)*	**Payoff recipient** *(days)*	**N. dictators**
NowNow	0	0	89
NowLater	0	35	83
LaterLater	35	35	90
LaterNow	35	0	88

Note: Treatment descriptions and number of dictators.

3.3. Intertemporal decision task for Self and for Other

After the modified Dictator Game, the dictators completed two individual tasks where they allocate money over time: in one task, they made decisions for themselves (i.e., Self), and in another separate task they decided on behalf of another participant who was assigned to the role of recipient in the Dictator Game (i.e., Other).[5] We elicit time preferences of each dictator for Self and for Other using a modified version of the so-called Convex Budget Method introduced by Andreoni and Sprenger (2012). The order of these tasks was counterbalanced between subjects. The only difference between the task for Self and the task for Other is that the allocation decisions affect either the earnings of the dictator or the earnings of a recipient, respectively. Each task consists of a list of 36 intertemporal allocation choices which were presented in a randomized order. In each decision, dictators are asked to allocate 100

[4] A time delay of 5 weeks is similar to the one used in a treatment comparable to our *LaterLater* in e.g., Kovarik (2009) or Dreber et al. (2016).

[5] For each dictator, the *Other* player was a randomly chosen participant who was assigned the role of recipient. Thus, the chance that a recipient is matched with the same dictator in both the Dictator Games and the Intertemporal Task is very small, but larger than zero.

tokens to a sooner and a later date. The sooner date t is 0, 14, 49 days after the experiment took place and the later date h is 35, 49, 63, 77, 84 and 112 days after the experiment took place. Each token allocated to the sooner date has a value a_t, and each token allocated to the later date has a value a_{t+h}. The value of each token allocated to the later date is fixed at a_{t+h} = €0.20, while the value of each token allocated to the sooner date a_t varies and takes the values a_t= €0.19, €0.18, €0.17, €0.16, €0.15 and €0.14. Dictators are informed that they can allocate any amount of tokens to one of the two dates. Table 3.3 shows the experimental parameters for the intertemporal tasks.

Table 3.3: **Intertemporal task. Experimental parameters.**

Set	t (sooner date)	h (delay)	1+*r*	Annual discount rates (%)
1	0	35	1.05; 1.11; 1.18; 1.25; 1.33; 1.43	54; 110; 170; 233; 301; 373
2	0	63	1,00; 1.05; 1.18; 1.25; 1.33; 1.54	0; 30; 94; 130; 167; 250
3	14	35	1.05; 1.11; 1.18; 1.25; 1.33; 1.43	54; 110; 170; 233; 301; 373
4	14	63	1,00; 1.05; 1.18; 1.25; 1.33; 1.54	0; 30; 94; 130; 167; 250
5	49	35	1.05; 1.11; 1.18; 1.25; 1.33; 1.43	54; 110; 170; 233; 301; 373
6	49	63	1,00; 1.05; 1.18; 1.25; 1.33; 1.54	0; 30; 94; 130; 167; 250

Note: $(1 + r)$ is the experimental gross interest rate, defined as in Andreoni and Sprenger (2012). $1 + r$ and the annual discount rates refer to each of the six intertemporal allocation choices within a set. In both tasks (i.e., Self and Other), the parameters are identical.

3.4. Procedure and payments

A total of 700 participants took part in the experiments. The 350 dictators were recruited using ORSEE (Greiner, 2015). The experiment was computerized and took place at the BEElab of Maastricht University - School of Business and Economics. The experiment was programmed with z-tree (Fischbacher, 2007).[6] We conducted 15 experimental sessions and the data were collected in March and October 2017. In 7 experimental sessions, we implemented the treatments *NowNow* and *LaterLater*. In the remaining 8 sessions, we collected data for the treatments *NowLater* and *LaterNow*. In each experimental session, approximately half of the participants were randomly allocated to one treatment implemented in that session. Each session took about 110 minutes, including the time participants spent reading the instructions at their own pace and a final questionnaire where they were asked to report demographic information and to answer questions regarding their spending attitudes and whether they believed the payment procedure. Around 80% of the dictators were students from the School of Business and Economics. The remaining 20% were from other disciplines such as psychology, biomedical sciences and law. Approximately 50% of the participants were male and the average age was 21 years.[7]

[6]The 350 recipients were recruited from the subject pool of BonnEconLab (Bonn, Germany) because of the limited number of participants registered at Maastricht University. Recipients completed a Qualtrics survey where they answered some hypothetical questions and provided their name, email address and IBAN. Recipients were told they would receive at least €4 for completing the survey and that additional earnings depended on the decisions of another participant in the experiment. Dictators were informed that their decisions will affect the earnings of another participant in the experiment, but they were not told that recipients were students of another university.

[7]Men and women are equally distributed across treatments ($\chi^2(3) = 2.41$, $p = 0.493$). Men in each treatment: *NowLater*: 57%; *LaterLater*: 47%; *NowNow*: 55%; *LaterNow*: 49%.

Each dictator completed the experiment in a cubicle isolated from other participants. In the Dictator Game, the decision situations (Table 3.1) were randomized at the individual level and each situation was displayed individually on the computer screen. In the experiment, dictators and recipients were labeled as A and B, respectively. In each decision situation of the Dictator Game, a dictator was randomly rematched with a new recipient. In the intertemporal tasks, the order of the tasks was counterbalanced in each experimental session and for each treatment: half of the participants completed first the intertemporal task for Self, and vice-versa.[8] When deciding on behalf of someone else, a dictator was matched with the same recipient for the whole duration of the task.

One randomly selected situation in each of the three tasks of the experiment was selected to determine the payments for dictator and the recipient. The earnings of the dictator therefore consists of the amount she decided to keep for herself in the selected situation in the Dictator Game plus the earning from the intertemporal choice task for Self. The earnings of recipients were determined by the dictators they were matched with and consisted of the amount given by the dictator in the Dictator Game task and of the allocation made by the dictator in the intertemporal choice task for Other. Participants earned €25.2 on average and all the earnings were paid via bank transfers.[9] Transfers were made on the day of the experiment making sure that participants received the earnings according to the decisions which were paid out. The possible payment dates were chosen to ensure that they will not be on the weekend and national holidays. In the Netherlands, bank transfers are usually received on the day of the transaction.

Because of the time delay associated with the payments, it is important to equalize transaction costs and credibility across treatments. To deal with this, we employed a similar approach as Andreoni and Sprenger (2012). First, all payments were transferred to the bank accounts that participants provided. Participants were truthfully informed that payments were issued by research assistants not otherwise involved in the experiment and that is was impossible to link their identity to their behavior. A second challenge was removing any convenience resulting from concentrating all payments in one period. To solve this, before starting each intertemporal task, dictators were told that they or the other participant would receive €1 sooner and €1 later, regardless of the allocation decision. All experimental earnings were added to this €2 baseline earning. Third, dictators were given a business card of the secretarial office of the Department of Economics of the School of Business and Economics and explicitly told to write an e-mail should they experience any problems with payments. Fourth, at the end of the experiment dictators were asked to write down the amounts they earned in each part of the experiment and the dates when the transfers will be made on two invoices. They were instructed to keep one invoice. Fifth, to minimize the cost of remembering the amounts they

[8]See Figures 3.7 and 3.8 in the Appendix for an example of the screen interface for the Dictator Game and Figure 3.9 and Figure 3.10 for the screen interfaces of the intertemporal tasks Self and Other, respectively.

[9]Bank accounts in the Netherlands and in Germany are linked to debit cards, which are widely used in most stores such as supermarkets, restaurants, and pubs without any transaction fees. The responses from the dictators in the post experimental questionnaire show that around 53% of the dictators pay more than 70% of their general expenditures by debit card and that 87% pay more than 30% of their general expenditures by debit card. 72% of the dictators prefer debit cards as a method of payment over cash and credit card.

earned, participants were told before starting the experiment that they would receive an individual email from research assistants not otherwise involved in the experiment with a summary of their earnings and the dates when the transfers will be made.[10]

4. Hypotheses

In this section, we describe the hypotheses. First, we focus on the effect of time delay on giving in the Dictator Game. Second, we study the relative price effect on giving. Finally, we analyze separately how giving depends on p_{keep} and p_{give}, and how different time delays affect this relationship. We base the hypotheses on the empirical literature reviewed in Section 2.

The first hypothesis refers to the effect of time delay on giving in the Dictator Game. The evidence in this domain is so far mixed. Therefore, we first focus on the analysis of the effect of time delay on giving aggregating over the different endowments and prices, and we propose the null hypothesis:

Hypothesis 1. *Aggregated over different endowments and prices, time delay does not affect giving.*

Based on standard microeconomic theory we assume that giving follows the "law of demand" and that only the relative price matters. We therefore propose the following hypotheses:

Hypothesis 2a. *At the aggregate level, giving decreases with increasing relative price of giving.* **2b.** *This relationship holds for all four treatments separately.*

Our design allows to investigate how time delay affects the sensitivity to different prices of keeping p_{keep} and prices of giving p_{give}. Even though according to standard microeconomic theory only the relative prices should be relevant, it is perceivable that it matters behaviorally if the absolute price of giving or the absolute price of keeping changes. We expect that giving decreases (increases) as the the price of giving (keeping) increases. We therefore propose the following exploratory null hypotheses:

Hypothesis 3a. *In absolute terms, the price sensitivity of giving is the same for the price of giving p_{give} as for the price of keeping p_{keep}.* **3b.** *This relationship holds for all the four treatments separately.*

5. Results

In this section, we first present the overall treatment (i.e., time delay) effect on giving (Hypotheses 1), and test whether this effect holds when taking into account the dictators' discount rates for Self and Other. We then study the relative price sensitivity and how this relationship is affected by time delays (Hypothesis 2), and whether it matters if the absolute price of giving or the absolute price of keeping changes (Hypothesis 3). We also discuss if previous results obtained in a setup similar to our *NowNow* treatment replicate in the

[10] Dictators highly trusted that the payments will be transferred (on a scale from 0 to 10, where 0 means do not trust at all and 10 means fully trust, the average was 9.27 and the median 10) and that it will be transferred on the designated day (average 8.84, median 10).

more complex setup we study. Thereafter, we explore whether the choices we observe are rationalizable and whether time delay affects the rationalizability of choices.

5.1. *Effects of time delay on giving*

We first look at the effect of time delay on giving at the treatment level aggregating over endowments and prices. Since endowments vary across decision situations, we define *giving* as the ratio of the number of tokens the dictator gives to the amount of tokens the dictator is endowed with.[11] Figure 3.1 shows giving in each of the four treatments.

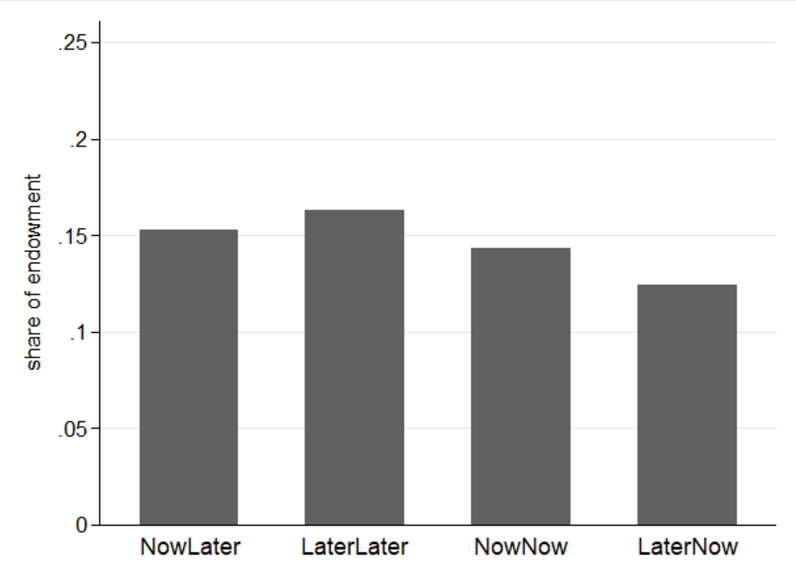


Figure 3.1: **Dictator Game. Mean of giving per treatment.**

Across all treatments, dictators share on average 14.6% of their endowments, and giving is similar across treatments (Table 3.4).[12] A Kruskal-Wallis test confirms the impression of no differences in giving across treatments ($\chi^2(3) = 2.07$, $p = 0.55$). Thus, when aggregating over all endowments and prices, time delay does not affect giving.

Column (1) in Table 3.5 shows the estimates of a random-effects Tobit regression including treatment dummy variables. The regression corroborates the result of the non-parametric test that none of the treatments has a different effect on giving compared to *NowNow*. Furthermore, a Wald test on the restriction that all coefficients of the treatments are jointly equal to zero does not reject this restriction ($p = 0.537$). Since time delay does not affect giving, we cannot reject Hypothesis 1.

Result 1 *At the aggregate level, time delay does not affect giving.*

[11] All the Wald tests presented are Holm corrected, unless otherwise stated. Regression tables using absolute giving as dependent variable are shown in Table 3.10 and 3.13 in the Appendix.

[12] The average amount of giving we observe is lower than the amount reported in the widely cited study by Engel (2011), where dictators allocate an average of 28% of the endowments. Note, however, that this average includes both students and non-students. Engel (2011) shows that on average, non-students give much more than students.

Table 3.4: **Effect of time delay on giving.**

Treatment	Mean	Std. Dev.	Median	N
NowLater	.153	.165	.089	83
LaterLater	.163	.170	.097	90
NowNow	.144	.156	.109	89
LaterNow	.124	.154	.073	88
Total	.146	.161	.083	350

Note: Mean, standard deviation, and median of giving in each treatment.

Table 3.5: **Dictator Game. Random-effects Tobit regression on giving across treatments.**

	(1)
	giving
NowLater	0.059
	(0.055)
LaterLater	0.037
	(0.054)
LaterNow	-0.013
	(0.054)
Constant	-0.051
	(0.039)
N	5250

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We use a random-effects Tobit regression with left and right censoring to account for the within-subject design and the censoring at zero and one. The dependent variable is giving coded as the ratio between the number of tokens shared with an anonymous participant over the amount of tokens endowed. NowNow is the omitted treatment.

We note that this result does not take potentially important heterogeneity in participants' discount rates into account. Therefore, we estimate these discount rates using the data of the intertemporal tasks and check whether *Result 1* is robust to taking these two aspects into account. For the estimation of both the discount rates for Self and Other, we follow the approach of Andreoni and Sprenger (2012) which requires additional assumptions. We thus assume a time separable constant relative risk aversion (CRRA) utility function discounted with a quasi-hyperbolic β-δ discounting function (e.g., Laibson, 1997).

$$U(c_t, c_{t+h}) = \frac{1}{\alpha}(c_t - \omega_1)^\alpha + \beta\delta^h \frac{1}{\alpha}(c_{t+h} - \omega_2)^\alpha, \tag{3.1}$$

where δ and β indicate the one period discount factor and the present bias parameter, respectively. The values c_t and c_{t+h} are the payoffs in the intertemporal choice task at the time t and $t + h$, respectively. α is the curvature parameter of the CRRA utility function. The terms ω_1 and ω_2 capture background consumption (i.e., wealth). The values of β, δ, and α in the intertemporal formulation of a Stone-Geary linear demand for c_t (see Equation 3.2) can be estimated using non-linear least squares. We do this separately for Self and Other. Following Andreoni and Sprenger (2012), we restrict $\omega_1 = \omega_2 = 0$ for the estimation of parameters at the individual level to limit the number of estimated parameters. The parameter estimates are reported in Table 3.11 in the Appendix. Overall, we do not observe a difference in discount rates for Self and Other (Wilcoxon signed-rank test, $p = 0.506$).[13] Several previous studies have found differences in discounting for Self and Other. However, the findings are inconsistent regarding the direction of the difference. While some studies have found that individuals are more impatient when they make decisions for themselves than for others (Albrecht et al., 2011; Shapiro, 2010) or a charity (Howard, 2013), others report that individuals choose more impatiently (de Oliveira and Jacobson, 2020) or equally patiently (Rong et al., 2019).

$$c_t = \left\{ \begin{array}{l} \left[\frac{1}{1+(1+r)(\beta\delta^h(1+r))^{\left(\frac{1}{\alpha-1}\right)}}\right]\omega_1 + \left[\frac{(\beta\delta^h(1+r))^{\left(\frac{1}{\alpha-1}\right)}}{1+(1+r)(\beta\delta^h(1+r))^{\left(\frac{1}{\alpha-1}\right)}}\right](m-\omega_2)\ if\ \ t=0 \\ \\ \left[\frac{1}{1+(1+r)(\delta^h(1+r))^{\left(\frac{1}{\alpha-1}\right)}}\right]\omega_1 + \left[\frac{(\delta^h(1+r))^{\left(\frac{1}{\alpha-1}\right)}}{1+(1+r)(\delta^h(1+r))^{\left(\frac{1}{\alpha-1}\right)}}\right](m-\omega_2)\ if\ \ t>0 \end{array} \right\} \tag{3.2}$$

We can now replicate our analysis of treatment differences in giving taking into account heterogeneity in discount rates for Self and Other. For this robustuness check, we split our sample into three groups depending on whether an individual's discount rate for Self is smaller, equal to, or higher than the discount rate for Other. We create the groups considering only the dictators for whom we could estimate the discount rates both for Self and Other (304/350). We classify the dictators in the three groups comparing their discount rate values. We then analyze whether there are differences in giving between treatments for each of the groups. In total, 114 dictators have a higher discount rate for Self than Other, for 66 dictators both rates are exactly equal, and for 124 dictators the discount rate for Self is lower than the one for Other. Also when looking at these groups separately, we do not find significant treatment differences in giving. Thus, we cannot reject Hypothesis 1 for any of the groups.[14]

5.2. *Relative price of giving*

In this section, we focus on the relative price of giving to the other participants, which we define as $k = p_{give}/p_{keep}$. Table 3.6 shows the estimates of two random-effects Tobit regressions investigating the effect of the relative price of giving k at the aggregate level and for each treatment, respectively. By looking at giving relative to the endowment, we control for variation in endowments.[15] Column (1) in Table 3.6 shows that the coefficient of the variable *Relative price k* is negative and significant ($p \leq 0.01$). This suggests that, at the aggregate level, giving decreases with increasing relative price of giving and we thus confirm Hypothesis 2a. Column (2) in Table 3.6 shows the

[13]We do not observe differences between Self and Other in β ($p = 0.212$), while α for Self is larger and significantly different from α for Other ($p \leq 0.001$).

[14]Table 3.12 and 3.13 in the Appendix show the regression analysis for each of these groups.

[15]See Table 3.10 in the Appendix for the regressions with absolute giving and endowment as control variable.

coefficients of the treatments dummy variables and their interactions with the *Relative price k*. The coefficient of *Relative price k* is negative and significant ($p \leq 0.001$), indicating that in *NowNow*, giving decreases with increasing relative price of giving. All the interaction coefficients are significant ($p \leq 0.001$), indicating that *Relative price k* affects giving differently with time delay compared to *NowNow*. In fact, Wald tests (i.e., *Relative price k* + *Treatment*×*Relative price k* = *0*) show that giving is insensitive to different prices of giving when there is time delay ($p \geq 0.131$).[16] We thus reject Hypothesis 2b.

Result 2 *Giving decreases with increasing relative price of giving only in NowNow.*

Table 3.6: **Dictator Game. Random-effects Tobit regression on giving across treatments. The effect of relative price *k*.**

	(1) giving	(2) giving
Relative price *k*	-0.009**	-0.058***
	(0.004)	(0.007)
NowLater		-0.024
		(0.056)
LaterLater		-0.055
		(0.055)
LaterNow		-0.081
		(0.056)
NowLater× Relative price *k*		0.065***
		(0.010)
LaterLater× Relative price *k*		0.071***
		(0.010)
LaterNow× Relative price *k*		0.054***
		(0.010)
Constant	-0.019	0.023
	(0.020)	(0.039)
N	5250	5250

Standard errors in parentheses

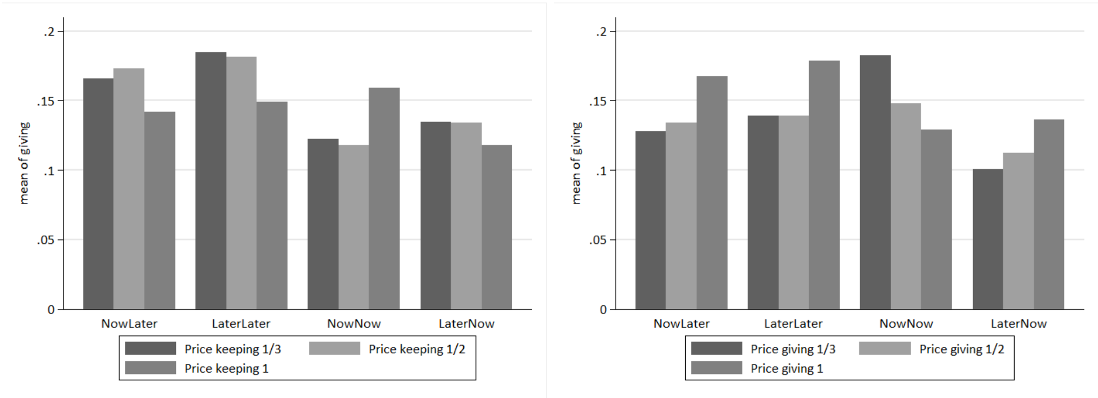
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We use a random-effects Tobit regression with left and right censoring to account for the within-subject design and the censoring at zero and one. The dependent variable is giving coded as the ratio between the number of tokens shared with an anonymous participant over the amount of tokens endowed. NowNow is the omitted treatment.

[16] p-values for each of the treatments: *NowLater*, $p = 0.556$; *LaterLater*, $p = 0.131$; *NowNow*, $p \leq 0.001$; *LaterNow*, $p = 0.629$.

5.3. *Absolute price sensitivity and time delay*

In this section, we take a closer look at the absolute price sensitivity. We analyze separately the effect of p_{keep} and p_{give} on giving across different time delays.



(a) p_{keep} (b) p_{give}

Figure 3.2: **Effect of price of keeping p_{keep} and price of giving p_{give} across time delays.**

Panel (a) of Figure 3.2 shows giving depending on the price of keeping p_{keep}, pooling across all values of p_{give}. Panel (b) shows giving depending on the price of giving p_{give}, pooling across all values of p_{keep}. Table 3.7 shows the coefficients of two random-effects Tobit regressions investigating the effect of p_{keep} and p_{give} at the aggregate level and for each treatment, respectively. Column (1) in Table 3.7 indicates that, at the aggregate level, the coefficient of p_{keep} is positive and significant ($p = 0.022$). This suggests that giving increases as p_{keep} increases. However, the results show that coefficient of the variable p_{give} is not significant. A test of $p_{keep} = -p_{give}$ indicates that at the aggregate level the price sensitivity of giving is the same for the price of keeping as for the price of giving ($p = 0.156$).[17] Thus, we partially reject Hypothesis 3a.

In Column (2) in Table 3.7, we investigate the relationship between p_{keep} and p_{give} and giving in different treatments. Column (2) includes the variables p_{keep} and p_{give} and their interactions with the treatments. A Wald test investigating whether $p_{keep} = -p_{give}$ for each of the treatments separately shows that in each treatment the price sensitivity of giving is the same for the price of keeping as for the price of giving (all p values ≥ 0.271).[18]

The coefficients of p_{keep} and p_{give} in *NowNow* are significantly positive and negative (both $p < 0.001$), respectively. This indicates that in the absence of time delay, dictators give more when it is relative more expensive to keep and keep more when it is relative more expensive to give. To further investigate the potential differences between *NowNow* and the other treatments, we first focus on the effect of p_{keep} on giving. The interaction coefficients between

[17]We note that this is the only result that changes when using the absolute giving as the dependent variable and controlling for the different endowment. In this case, we still observe that at the aggregate level p_{give} is not statistically significant, but the sensitivity of giving is significantly different between the p_{keep} and p_{give} ($p = 0.065$) (See Table 3.10 in the Appendix).

[18]E.g., for *NowLater*, we test if $p_{keep} + \textit{NowLater} \times p_{keep} = -(p_{give} + \textit{NowLater} \times p_{give})$.

the treatment variables and p_{keep} indicate that the effect of p_{keep} on giving in *NowNow* is significantly different from the effect in *NowLater* ($p = 0.015$) and in *LaterLater* ($p = 0.002$), but it is not different from the effect in *LaterNow* ($p = 0.159$). To test the direction of the effect of p_{keep}, we conduct a Wald test that indicates that p_{keep} has a positive effect on giving in *NowNow* ($p \leq 0.001$), but the effect is not different from zero in the other treatments (*NowLater*, $p = 0.800$; *LaterLater*, $p = 0.984$; *LaterNow*, $p = 0.289$).[19]

We now turn our attention to the effect of p_{give} on giving. The interaction terms between p_{give} and the treatment variables are all significant (all $p \leq 0.001$), indicating that p_{give} has a different effect in all the other treatments compared to *NowNow*. The Wald test we use to test the direction of the effect of p_{give} indicates that p_{give} has a negative effect on giving in *NowNow* ($p \leq 0.001$), but the effect is not different from zero in the other treatments (*NowLater*, $p = 0.235$; *LaterLater*, $p = 0.271$; *LaterNow*, $p = 0.303$). We thus reject Hypothesis 3b and propose our third result:

Result 3 *Only in NowNow, do dictators give more (less) when keeping (giving) is more costly. In all treatments involving time delay, dictators are insensitive to changes in the prices.*

[19]To test for the direction of the effect of p_{keep} on giving, we use the test: $p_{keep} + NowLater \times p_{keep} = 0$; $p_{keep} + LaterLater \times p_{keep} = 0$; $p_{keep} + LaterNow \times p_{keep} = 0$, and $p_{keep} = 0$. For testing the direction of the effect of p_{give} on giving we use the same approach, replacing p_{keep} with p_{give}.

Table 3.7: **Dictator Game. Random-effects Tobit regression on giving across treatments. The effect of p_{keep} and p_{give}.**

	(1) giving	(2) giving
p_{keep}	0.036*	0.116***
	(0.016)	(0.032)
p_{give}	0.004	-0.114***
	(0.015)	(0.031)
NowLater		0.025
		(0.082)
LaterLater		0.036
		(0.081)
LaterNow		-0.083
		(0.082)
NowLater $\times$ p_{keep}		-0.108*
		(0.044)
LaterLater$\times$ p_{keep}		-0.137**
		(0.044)
LaterNow$\times$ p_{keep}		-0.063
		(0.045)
NowLater$\times$ p_{give}		0.161***
		(0.043)
LaterLater$\times$ p_{give}		0.147***
		(0.043)
LaterNow$\times$ p_{give}		0.164***
		(0.044)
Constant	-0.062*	-0.058
	(0.029)	(0.058)
N	5250	5250

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We use a random-effects Tobit regression with left and right censoring to account for the within-subject design and the censoring at zero and one. The dependent variable is giving coded as the ratio between the number of tokens shared with an anonymous participant over the amount of tokens endowed. NowNow is the omitted treatment.

5.4. Comparison with previous findings in the literature

The data in Andreoni and Miller (2002) refer to a setup without time delay (i.e., our *NowNow*) and suggest that the amount of giving decreases in the price of giving and increases in the price of keeping, although this is not formally tested. However, *NowNow* seems to be special in this respect, as the dictators are insensitive to prices in the treatments involving a time delay.[20]

Looking at giving decisions for different p_{give} and p_{keep} separately also allows us to compare our results to some previous findings in the literature. In the "classic" Dictator Game (*NowNow*, p_{give} and $p_{keep} = 1$), we find that average giving is 14.8%. This is in line with the results documented in the review by Camerer (2003) where dictators also allocate between 10% and 25%. The amount we find is lower than what is reported in Andreoni and Miller (2002) and Fisman et al. (2007) where dictators allocate approximately 23% and 19% of the endowment to the recipient respectively. These differences are likely due to the fact that in our experiment the role of participants was kept fixed in all allocation situations whereas participants switch roles in Andreoni and Miller (2002) and Fisman et al. (2007). By eliminating potential perceived reciprocity, keeping roles fixed could lead to lower giving (Iriberri and Rey-Biel, 2011).

Two previous incentivized studies (Kovarik, 2009; Dreber et al., 2016) compare giving in a Dictator Game when the payment for both the dictator and the recipient occur immediately relative to a situation in which both are paid at a later point in time (*NowNow* vs. *LaterLater*). Both studies find that giving is lower when payments are postponed for both the dictator and the recipient. In Kovarik (2009), where p_{give} and $p_{keep} = 1$, average giving in *NowNow* lies around 15% of the endowment, while giving decreases to around 5% of the endowment when the decision is implemented 22 days from the experiment.[21] In Dreber et al. (2016), where $p_{give} = 0.5$ and $p_{keep} = 1$ and the experiment is conducted on the online platform Amazon Mechanical Turk, dictators give on average 38% of their endowment in *NowNow*, while the average giving is around 33% of the endowment when the payments are delayed by 30 days.[22] In our experiment, we do not replicate these findings. When p_{give} and $p_{keep} = 1$, average giving in *NowNow* is 14.8% of the endowment and 17% in *LaterLater* (Wilcoxon rank-sum test, $p = 0.390$). When $p_{give} = 0.5$ and $p_{keep} = 1$, giving is 14.8% in *NowNow* and 13.9% in *LaterLater* (Wilcoxon rank-sum test, $p = 0.872$).

Gender differences in sensitivity to the relative price of giving

In a widely cited paper, Andreoni and Vesterlund (2001) find that the demand curves for altruism for men and women cross. That is, when giving is expensive (i.e., $k \geq 1$), women are more generous whereas when giving is cheaper (i.e., $k < 1$), men are more generous.[23] Their setup is equivalent to

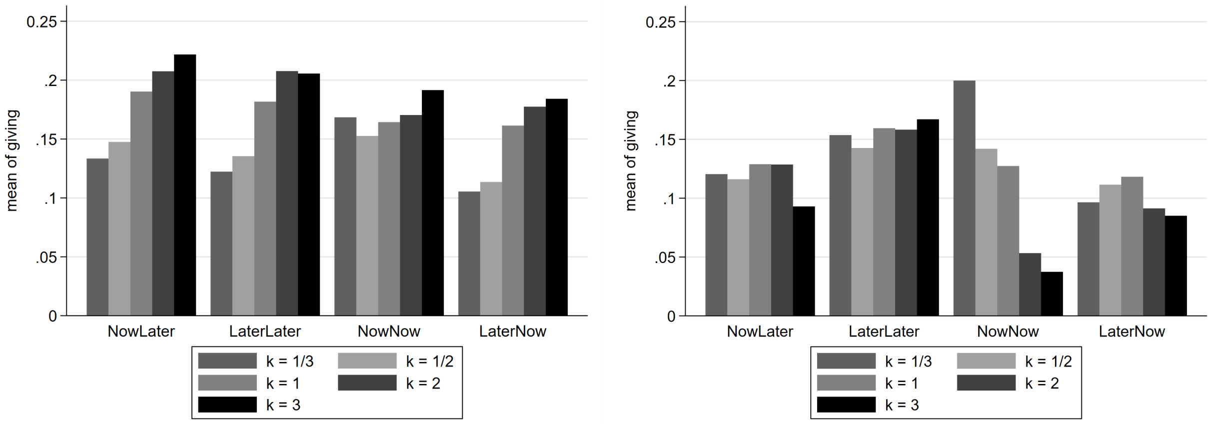
[20] However, we note again that the effect of p_{keep} is not not different between *NowNow* and *LaterNow*.

[21] Kovarik (2009) does not report the exact values and these numbers are inferred from Figure 1 in his paper. Kovarik (2009) includes 24 observations for t = 0, and 30 observations for the t = 22.

[22] In a footnote, Dreber et al. (2016) mention that in an initial pilot experiment with a student sample, dictators kept 90% of their endowment. In contrast, the equivalent number for the MTurk experiment was 62%.

[23] See also Croson and Gneezy (2009), Visser and Roelofs (2011), Boschini et al. (2012) for similar results.

our *NowNow* treatment. In the following analysis, we test whether this finding replicates in settings that involve time delays.



(a) Women, N = 181 (b) Men, N = 169

Figure 3.3: **Gender differences in giving across treatments for each relative price of giving (k).**

Figure 3.6 shows the differences in the sensitivity of giving to the relative price of giving *k* for each treatment and both genders. A first glance at the figure already suggests that men and women differ in their sensitivity to the relative price of giving *k* in settings which involve time delays. At the aggregate level, we observe that women are more generous than men (Wilcoxon rank-sum test, $p = 0.004$). The finding that women are more generous than men when giving is expensive ($k \geq 1$) seems to hold independently of time delays (Wilcoxon rank-sum tests for giving when $k \geq 1$, all treatments $p < 0.001$). However, the finding that men are more generous when giving is cheap ($k < 1$), does not generalize to all treatments. When giving is cheap, men give more than women in *LaterLater* and in *NowNow* (Wilcoxon rank-sum tests for giving when $k < 1$, $p \leq 0.001$), while women give more than men in *NowLater* and in *LaterNow* (Wilcoxon rank-sum tests for giving when $k < 1$, $p \leq 0.001$). The demand curves for altruism therefore do not cross in the latter two treatments.

To look at the gender differences more formally, we run a random-effects Tobit regression for each treatment using giving as dependent variable. We include a dummy for male participants, the relative price of giving k, and the interaction term between male and the relative price k (Table 3.8). In all treatments, we replicate the higher sensitivity of men to the relative price of giving found by Andreoni and Vesterlund (2001)—the interaction term of male and k is negative and significant (all $p \leq 0.012$). The coefficient of Relative price k is positive and significant in all treatments except in *NowNow* (in all treatments, $p \leq 0.001$; $NowNow$, $p = 0.970$). This suggests that, in the presence of time delays, women give more as giving becomes more costly. A Wald test on the restriction *Relative price k* + *Male* $\times$ *Relative price k* $= 0$, indicates that men tend to be efficiency oriented and give less when giving becomes more costly in all treatments, except $LaterLater$ (in all treatments, $p \leq 0.001$; $LaterLater$, $p = 0.685$)

In sum, our results for *NowNow* tend to replicate the findings of Andreoni

Table 3.8: **Dictator Game. Effect of relative price of giving across treatments, by genders.**

	(1) NowLater	(2) LaterLater	(3) NowNow	(4) LaterNow
Male	-0.044	-0.050	0.071	-0.004
	(0.066)	(0.084)	(0.094)	(0.065)
Relative price k	0.034***	0.030**	0.000	0.031***
	(0.007)	(0.010)	(0.011)	(0.008)
Male × Relative price k	-0.063***	-0.034*	-0.208***	-0.067***
	(0.011)	(0.014)	(0.021)	(0.012)
Constant	0.048	-0.011	-0.030	-0.028
	(0.043)	(0.060)	(0.063)	(0.048)
Observations	1245	1350	1335	1320

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We use a random-effects Tobit regression with left and right censoring to account for the within-subjects design and the censoring at zero and at one. The dependent variable is the fraction of income passed. We include dummy variables for male participants and the interaction term with relative price of giving k.

and Vesterlund (2001). The finding that women are more generous than men when giving is expensive ($k \geq 1$) also holds with time delay. However, when giving is cheap ($k < 1$), giving is not higher for men in all treatments.[24]

5.5. Are choices rationalizable?

Our experimental design allows us not only to capture the potential effect of delay on giving but also to investigate if individual choices are rationalizable. That is, whether they can be represented by a utility function. This step is important for both a better understanding of behavior and for a more accurate modelling of decision making. We extend the analysis of Andreoni and Miller (2002) to time-delay treatments. Given the additional layer of complexity, we might expect a lower degree of rationalizability in treatments which involve time delays.

Note that in the Netherlands, the interest rate over 5 weeks in 2017 was very close to zero.[25] Therefore, the interest rates in our budget constraints for those treatments that involve payments made at a later point in time are set to zero. We test the consistency with utility maximization using the Generalized Axioms of Revealed Preference (GARP).[26] Since we have linear budget constraints, satisfying GARP is both a necessary and sufficient condition for having well-behaved preferences (Varian, 1982).

One way to measure the severity of GARP violations is the Critical Cost Efficiency Index (CCEI) (Afriat, 1972), which measures the amount by which

[24]See Figure 3.6 in the Appendix for a graph similar to the one in Andreoni and Vesterlund (2001) that shows the payoff passed as a fraction of income defined as $\frac{\text{Tokens passed} \times \text{Pass value}}{\text{Endowment} \times \text{Hold value}}$, on the x-axis and the relative price of giving k on the y-axis.

[25]The interest rate on the main refinancing operations of the ECB was 0% during 2017.

[26]GARP requires that if an option A is directly revealed preferred to an option B, then A is not strictly within the budget set when B is chosen.

budget sets should be relaxed to avoid GARP violations. The CCEI lies between zero and one and the closer it is to one, the smaller the severity of GARP violations. Table 3.9 lists for each treatment the number of subjects who violate GARP at least once as well as the mean and median of the CCEI. Across treatments, between 35.6% and 41% of the subjects violate GARP at least once. In the literature, GARP violations have only been studied in contexts similar to the *NowNow* treatment. In *NowNow*, we observe that 37.1% of subjects violate GARP at least once and that 13.5% have a CCEI<0.95. These values are comparable to the ones reported in Andreoni and Miller (2002). It should be noted, however, that in our experiment subjects have many more opportunities to violate GARP since they have to make choices for a larger menu of budget sets. We do not observe a statically significant difference across treatments in GARP violations ($\chi^2(3) = 0.645$, $p = 0.886$) as well as the CCEI (Kruskal-Wallis test, $p = 0.772$) suggesting that time delay does not affect the degree of rationality in giving.

Table 3.9: **Dictator game. GARP violations and CCEI per treatment**

	NowLater	LaterLater	NowNow	LaterNow
# (%) subjects who violates GARP	34 (41.0%)	32 (35.6%)	33 (37.1%)	31 (35.2%)
Mean (median) of CCEI	0.994 (1)	0.987 (1)	0.977 (1)	0.989 (1)
N	83	90	89	88

Next, we generate a benchmark level of consistency with which we may compare our CCEI values. Following Andreoni and Miller (2002), we employ the test designed by Bronars (1987) that uses the choices of a hypothetical subject who randomizes uniformly among all allocations on each budget line as a benchmark. We calculate the CCEI generated by a sample of 25,000 of such hypothetical subjects and this hypothetical distribution is significantly different from our actual data in each of the treatments (Kolmogorov-Smirnov test, p values < 0.001).[27] It is clear that a significant majority of subjects came much closer to consistency with utility maximization than random hypothetical agents. We therefore conclude that most subjects exhibit behavior that appears to be rationalizable in the sense that their choices nearly satisfy GARP, so that the violations are minor enough for the purposes of recovering preferences or constructing appropriate utility functions also when time delay is involved.

6. Conclusion

In many contexts, prosocial behavior has an intertemporal dimension. From the perspective of the experimental literature, an important question regarding altruism is whether the findings of the substantial literature on Dictator Games in a setting without delayed payments carry over to settings in which payments are delayed. This paper provides a comprehensive exploratory analysis of giving in Dictator Games with different time delays, and our findings help to close this knowledge gap. In addition, we consider how giving is affected by different prices of keeping and of giving, and how this relationship depends on time delay.

In several aspects, the behavior observed in the treatments with delayed

[27]These two distributions of CCEI are shown in Figure 3.5 in the Appendix.

payments is not different from that in *NowNow*. We find no differences in aggregate giving between the four treatments and this result holds also when taking into account the differences in discount rates for own and the recipient's payoffs. In all four treatments, men react differently then women to the price of giving. We do find, however, also important differences. Dictators react to different prices of giving and keeping only in *NowNow*, where dictators give more when keeping is costly and give less when giving is costly. In the other treatments which involve delayed payments, dictators are insensitive to changes in prices. From a more applied perspective, campaigns relying on matching donations to encourage giving could thus be less efficient if they involve long time delays. Regarding gender, we provide evidence that the finding that men give more than women when giving is cheap (Andreoni and Vesterlund, 2001) does not replicate in two of the treatments involving delayed payments.

The insensitivity to different prices in treatments that involve delayed payments compared to *NowNow* does not seem to be driven by different discount rates for self and other. In contrast to some findings in the literature (e.g., Albrecht et al., 2011; Shapiro, 2010; Howard, 2013; de Oliveira and Jacobson, 2020), we do not observe that dictators discount the utility received from own payoff and the utility received from the other's payoff at different rates. Also the added layer of complexity in the treatments which involve time delay does not seem the cause of this different sensitivity to prices. Dictators' choices satisfy GARP in all the four treatments and for most subjects it appears possible to rationalize their behavior by well-behaved utility function. Anonymity of choices and reputational concerns were also kept constant across treatments.

Overall, our findings contribute to the growing literature that analyzes how time delay affect social preferences. The picture that emerges from this exploratory investigation can inform theoretical frameworks that aim to model intertemporal social preferences (e.g., Dreber et al., 2016; Andreoni and Serra-Garcia, 2020). Future research could focus on identifying the channels that explain why the sensitivity to different prices in a treatment similar to our *NowNow* differs from situation where there is a time delay. For instance, one possible avenue could be design and experiment with more variation in time delays for both self and other to test whether the decline in sensitivity to prices occur as long as there is a short delay.

7. Appendix

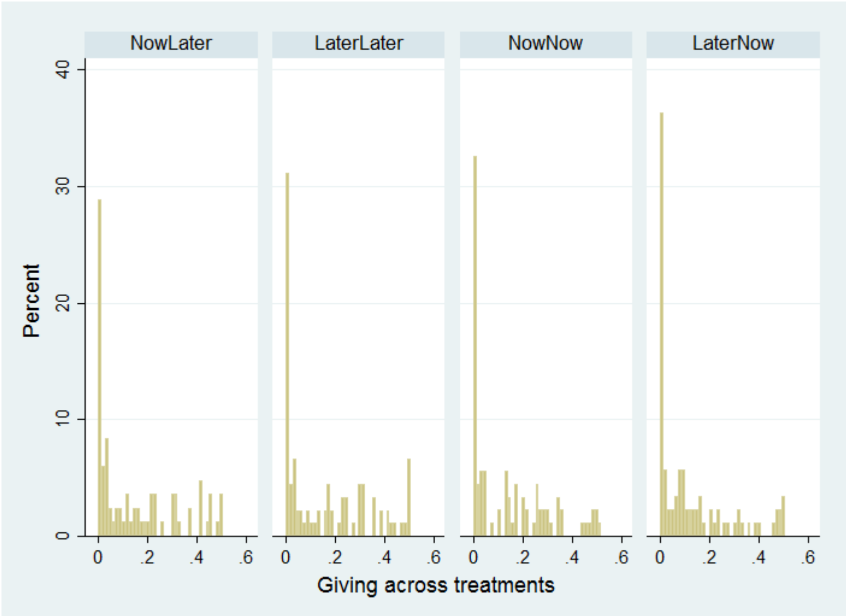


Figure 3.4: **Dictator game. Giving across treatments.**

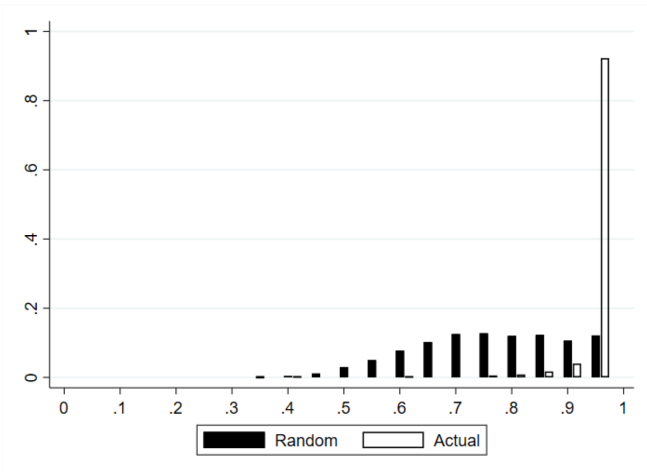


Figure 3.5: **CCEI distributions.** CCEI generated by a sample of 25,000 hypothetical subjects and actual experimental data. Note that the hypothetical subjects randomize uniformly among all allocations on each budget line as a benchmark.

Table 3.10: **Dictator Game. Tobit regression using absolute giving as dependent variable.**

	(1) giving	(2) giving	(3) giving	(4) giving	(5) giving
Nowlater	4.334		-0.996		4.024
	(3.869)		(3.974)		(5.761)
LaterLater	2.716		-2.865		3.143
	(3.806)		(3.912)		(5.708)
LaterNow	-0.749		-5.007		-2.425
	(3.830)		(3.936)		(5.770)
Endowment	0.222^{***}	0.218^{***}	0.218^{***}	0.210^{***}	0.210^{***}
	(0.012)	(0.012)	(0.012)	(0.014)	(0.014)
Relative price *k*		-0.779^{**}	-3.870^{***}		
		(0.250)	(0.514)		
NowLater× Relative price *k*			4.251^{***}		
			(0.702)		
LaterLater× Relative price *k*			4.443^{***}		
			(0.699)		
LaterNow× Relative price *k*			3.437^{***}		
			(0.713)		
p_{keep}				3.387^{**}	9.297^{***}
				(1.225)	(2.306)
p_{give}				0.754	-5.881^{**}
				(1.204)	(2.198)
NowLater× p_{keep}					-8.255^{**}
					(3.097)
LaterLater× p_{keep}					-8.924^{**}
					(3.087)
LaterNow× p_{keep}					-5.722
					(3.139)
NowLater× p_{give}					9.231^{**}
					(2.997)
LaterLater× p_{give}					8.960^{**}
					(2.985)
LaterNow× p_{give}					8.429^{**}
					(3.019)
Constant	-19.557^{***}	-16.694^{***}	-14.406^{***}	-20.414^{***}	-21.706^{***}
	(2.870)	(1.712)	(2.945)	(2.074)	(4.109)
N	5250	5250	5250	5250	5250

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We use a random-effects Tobit regression with left censoring to account for the within-subjects design and the censoring at zero. The dependent variable is absolute giving. NowNow is the omitted treatment.

A.1. Intertemporal tasks

Table 3.11: **Intertemporal Task**. Descriptive statistics of annual interest rate, present bias and curvature parameters for Self and Other aggregating across treatment, and separately for each treatment. All participants for whom we could estimate parameters are included.

Variable	N	Median	5th Percentile	95th Percentile	Min	Max
Interest rate Self	326	0.372	0.000	10.546	-1.000	4358.105
Interest rate Other	310	0.468	-0.013	9.206	-1.000	5.4e+35
Present Bias β Self	326	1.000	0.806	1.115	0.000	1678.386
Present Bias β Other	310	1.000	0.886	1.183	0.096	5.0e+04
CRRA Curvature: α Self	326	0.998	0.737	0.999	0.000	41.624
CRRA Curvature: α Other	310	0.986	0.733	0.999	0.000	1.000
Interest rate Self (NL)	79	0.842	0.054	18.339	-0.322	4358.105
Interest rate Self (LL)	82	0.226	0.000	7.628	-0.996	66.448
Interest rate Self (NN)	81	0.346	0.161	10.395	-0.200	39.925
Interest rate Self (LN)	84	0.382	-0.068	18.010	-1.000	697.990
Interest rate Other (NL)	73	0.469	0.000	6.452	-0.748	25.933
Interest rate Other (LL)	78	0.585	-0.013	3.882	-0.964	44.237
Interest rate Other (NN)	80	0.170	-0.115	8.349	-1.000	147.127
Interest rate Other (LN)	79	0.706	-0.119	39.532	-0.924	5.4e+35
Present Bias β Self (NL)	79	1.000	0.333	1.154	0.061	1.434
Present Bias β Self (LL)	82	1.000	0.806	1.069	0.000	2.016
Present Bias β Self (NN)	81	1.000	0.853	1.072	0.378	1.386
Present Bias β Self (LN)	84	1.000	0.893	1.119	0.082	1678.386
Present Bias β Other (NL)	73	1.000	0.888	1.140	0.804	2.428
Present Bias β Other (LL)	78	1.000	0.837	1.231	0.096	5.329
Present Bias β Other (NN)	80	1.000	0.852	1.253	0.301	91.126
Present Bias β Other (LN)	79	1.000	0.924	1.152	0.828	5.0e+04
CRRA Curvature: α Self (NL)	79	0.988	0.656	1.000	0.126	1.000
CRRA Curvature: α Self (LL)	82	0.998	0.737	0.999	0.000	1.000
CRRA Curvature: α Self (NN)	81	0.999	0.830	0.999	0.672	1.000
CRRA Curvature: α Self (LN)	84	0.996	0.621	0.999	0.422	41.624
CRRA Curvature: α Other (NL)	73	0.975	0.696	0.999	0.000	1.000
CRRA Curvature: α Other (LL)	78	0.988	0.737	0.999	0.003	1.000
CRRA Curvature: α Other (NN)	80	0.998	0.669	0.999	0.211	1.000
CRRA Curvature: α Other (LN)	79	0.975	0.827	0.999	0.538	1.000

A.2. *Treatment difference by subgroups depending on the discount rates*

Considering the dictators for whom we could estimate the discount rates both for Self and Other, 114 dictators have a higher discount rate for Self than Other, for 66 dictators both rates are equal, and for 124 dictators the discount rate for Self is lower than for Other. Table 3.12 show the effect of time delay on giving for each of this subgroups. In none of the subgroups we find significant treatment differences in giving. A Wald test at the subgroup level on the restriction that all coefficients of the treatments are jointly equal to zero does not reject this restriction (all $p \geq 0.465$). Table 3.13 shows the same analysis using absolute giving as dependent variable. Also in that case, we find that in none of the subgroups we observe significant treatment differences in giving (all $p \geq 0.475$).

Table 3.12: **Dictator Game. Tobit regression on giving across treatments, by subgroups of discount rate Self vs. Other.**

	(1) Self=Other	(2) Self<Other	(3) Self>Other
NowLater	0.039	0.035	0.023
	(0.138)	(0.090)	(0.092)
LaterLater	0.028	-0.014	0.087
	(0.136)	(0.085)	(0.097)
LaterNow	-0.100	-0.026	-0.074
	(0.164)	(0.082)	(0.095)
Constant	-0.139	0.056	-0.071
	(0.093)	(0.067)	(0.063)
Observations	990	1860	1710

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We use a random-effects Tobit regression with left and right censoring to account for the within-subjects design and the censoring at zero and one. The dependent variable is giving coded as the ratio between the number of tokens shared with an anonymous participant over the amount of tokens endowed. NowNow is the omitted treatment.

Table 3.13: **Dictator Game. Tobit regression with absolute giving as dependent variable, by subgroups of discount rate Self vs. Other.**

	(1) Self=Other	(2) Self<Other	(3) Self>Other
NowLater	3.014	2.577	1.685
	(9.791)	(6.294)	(6.576)
LaterLater	2.742	-0.963	6.129
	(9.626)	(5.929)	(6.930)
LaterNow	-6.729	-1.668	-5.311
	(11.583)	(5.749)	(6.791)
Endowment	0.241***	0.213***	0.228***
	(0.033)	(0.019)	(0.022)
Constant	-27.256***	-11.201*	-21.469***
	(7.060)	(4.859)	(4.809)
Observations	990	1860	1710

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We use a random-effects Tobit regression with left censoring to account for the within-subjects design and the censoring at zero. The dependent variable is the absolute giving. NowNow is the omitted treatment.

A.3. Gender differences using fraction of income passed

Here we follow the approach of Andreoni and Vesterlund (2001) and in Figure 3.6 we show the differences in the sensitivity of giving to the relative price of giving k for each treatment and both genders. The figure describes the fraction of the value of the dictator's endowment that is passed to the receiver, that is $\frac{\text{Tokens passed} \times \text{Pass value}}{\text{Endowment} \times \text{Hold value}}$, on the x-axis and the relative price of giving k on the y-axis. The pattern shows that the payoff passed by men and women decreases as the relative price of giving k increases, and that the fraction of income passed by men and women differ in their sensitivity to the relative price.

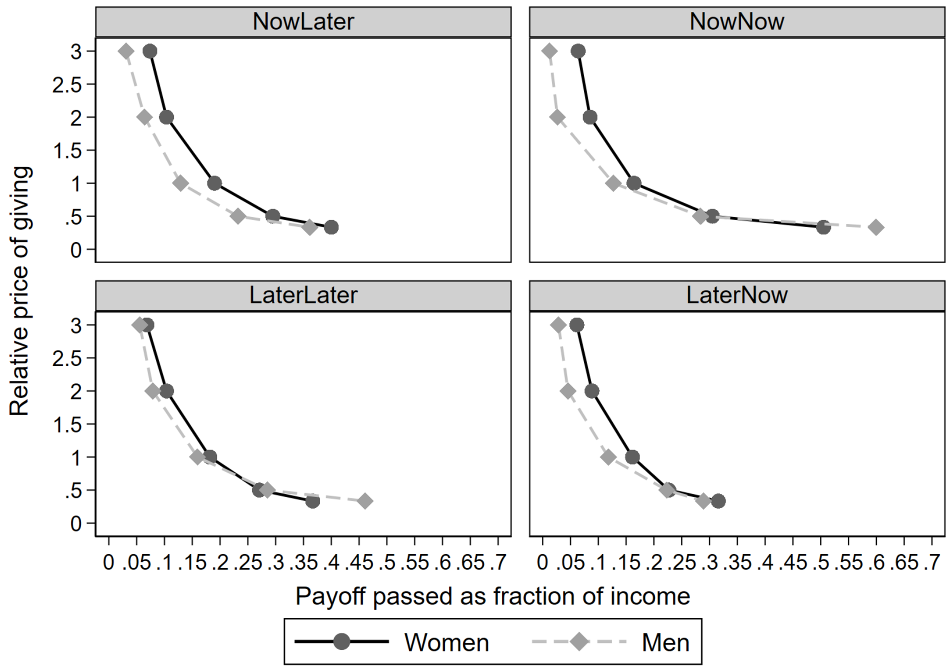


Figure 3.6: **Dictator Game. Payoff passed as fraction of income. Gender differences across treatments for each relative price of giving.**

7.4. *Instructions Dictator Game*

Welcome to this experiment! [28]
In this experiment, each participant will be either in the role of **PLAYER A** or **PLAYER B**. Throughout the entire experiment **your role will not change**. This experiment consists of several parts. You will receive information on each part immediately before the part starts. After you finish these parts, we also ask you to complete a questionnaire. Before we begin, we need to remind you that the earnings from this experiment will be paid by bank transfer, and you will be requested to provide a bank account number on your name in a later stage of this experiment. The payment dates are exactly as described in the decisions. We will send you today an email with a summary of your earnings and payment dates. On your desk there is a business card with the contact details of the Secretarial Office AE1 of the School of Business and Economics. Please keep this in a safe place. Please get in touch with the Secretarial Office should there be any problems with the payments. During and after the experiment, your identity will remain unknown to all other participants. We ask you not to communicate with other people during the experiment. If, at any stage, you have a question, please raise your hand. An experimenter will come to answer your question in private.

Your personal information

At the end of the experiment we will need to collect the following information from you: name, bank account number on your name, and email address. Your personal information will only be seen by a third person not involved in the experiment and in the subsequent data analysis and it will be used only for the purpose of payments. It will therefore be impossible to connect your identity to your earnings and behavior. Your anonymity is thus guaranteed. You have been randomly assigned the role of **PLAYER A**. Throughout the entire experiment **your role will not change**.

Detailed instructions - PART 1

In PART 1 you are asked to make a series of **15** decisions about how to divide a set of **tokens** between yourself and another participant who has been assigned the role of PLAYER B. You and the other participant will be paired randomly and you will not be told each other's identity. In each decision situation you will be rematched randomly with one participant who is in the role of PLAYER B. It is very unlikely that you will be matched with the same PLAYER B more than once.
As you divide the tokens, you and the other participant will **earn points**. These points will be converted into Euro at the rate of 1 point = 0.10 €. For example, if you earn 58 points this would be worth 5.80 € in this part of the experiment.

Both your earnings and PLAYER B's earnings in this part will be transferred in 5 weeks from today.[29]

Figure 3.7 below represents an example of a decision situation you may face during PART 1. *Note:* In the example below 'today' is the 9th of August 2017. In the experiment 'today' will refer to the real date of today.

- The information related to **YOU** will be displayed in **Brown**.
  The information related to **PLAYER B** will be displayed in **Blue**.
- On the top of the screen you see a calendar. The colored date in the calendar indicates when a bank transfer with the earned money will be made to **YOU**

[28] This instructions are an adapted version of the instructions used for the treatment *LaterLater*

[29] This was rephrased in each treatments. In $NowLater$: In this part, your earnings will be transferred today and PLAYER B's earnings will be transferred in 5 weeks from today. In $LaterNow$: In this part, your earnings will be transferred in 5 weeks from today and PLAYER B's earnings will be transferred today. In $NowNow$:Both your earnings and PLAYER B's earnings in this part will be transferred today.

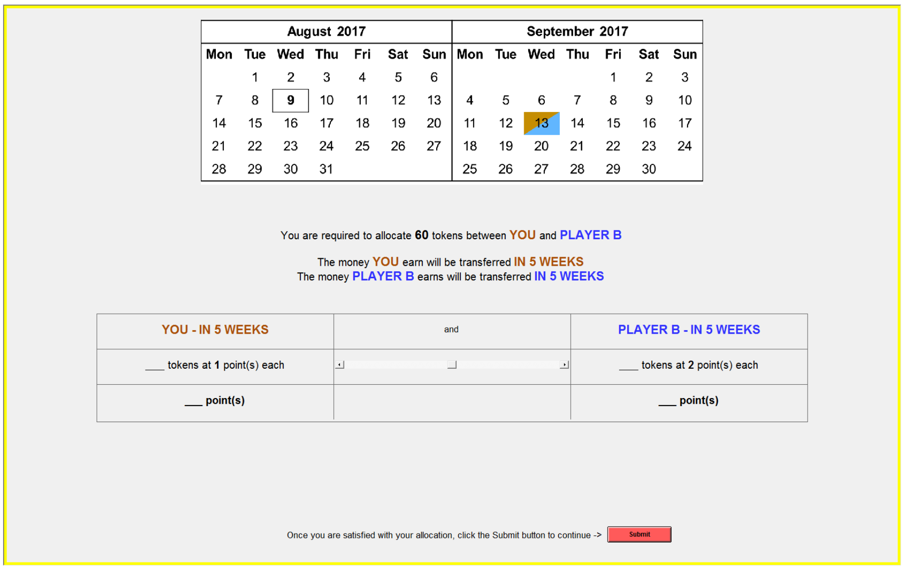


Figure 3.7: Dictator game - Example 1

and to **PLAYER B**. The date when the bank transfer will be made to **YOU** is highlighted in **Brown**. The date when the bank transfer will be made to **PLAYER B** is highlighted in **Blue**. The date of today is always **bold** and surrounded by a square shape (in the example, today is the 9th of August 2017). In the example shown in Figure 3.7, the bank transfers to **YOU** and **PLAYER B** will be made in 5 weeks from today (in the example 13th of September).

- Below the calendar you see the amount of tokens you are asked to divide. In the example, you are given a total of **60 tokens**. Note that in the experiment the amount of tokens will vary across decision situations.

- In the lower part of the screen you see a table. The column in the middle of the table displays a slider you should use to allocate tokens between **YOU** and **PLAYER B**. To make your allocation decision, you have to either scroll the bar or click the arrows. On the left column of the table you will see the amount of tokens you are allocating to **YOU** and your corresponding earnings in points for this decision situation. The amount of tokens you are allocating to **PLAYER B** and his/her total earning in points will be displayed in the right column of the table. Please note that you can keep all the tokens, keep some and pass some, or pass all the tokens to **PLAYER B**.

Note: *In different decision situations, tokens kept and tokens passed will have different earnings consequences in points.*

In this example, **YOU** will receive **1 point** for every token you hold, and **PLAYER B** will receive **2 points** for every tokens you pass. For instance, if you hold 60 and pass 0 tokens, **YOU** will receive 60 points and **PLAYER B** will receive no points. If you hold 0 tokens and pass 60, **YOU** will receive no points and **PLAYER B** will receive 60 x 2 = 120 points. Note that you could choose any number between 0 and 60 tokens to hold. For instance, you could choose to hold 29 tokens and pass 31. In this case **YOU** would earn 29 x 1 = 29 points and **PLAYER B** would receive 31 x 2 = 62 points. The points **YOU** earn will be displayed on the left column. The points **PLAYER B** earns will be displayed on the right column of the table.

Note: *Points will be exchanged to money at the rate of 1 point = 0.10 €.*
Here is another example:

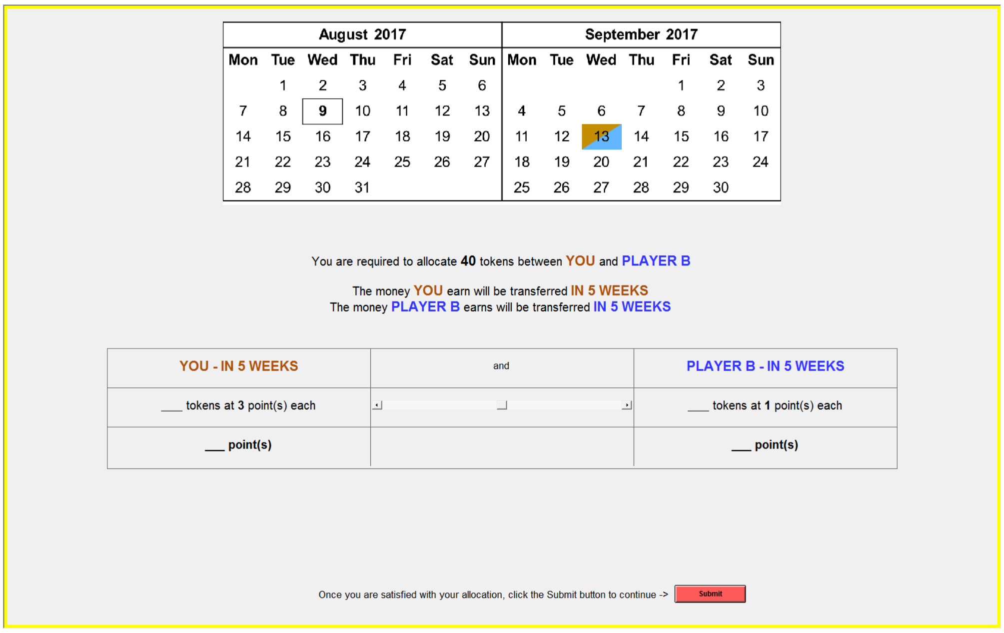


Figure 3.8: Dictator game - Example 2

In this example you are given a total of **40 tokens**. Every token you keep earns **YOU** 3 points, and every token you pass earns **PLAYER B** 1 point. You are free to choose any position of the cursor. Before clicking on the **Submit** button on the right bottom of the screen, you can change your decision as many times as you wish. After you have confirmed your decision, changes are not possible anymore and you will move to the next decision situation.

Important: You will make decisions in a number of different decision situations. Consider each decision situation carefully because they may **differ** in the **amount of tokens** you are asked to divide and in the **value** of each token you allocate to **YOU** and to **PLAYER B**. All decision situations are completely independent of each other and in each decision situation you will be randomly rematched with one participant who is in the role of **PLAYER B**. Therefore, you should consider each choice independently and carefully.

Determination of Earnings - PART 1

- **PLAYER B** will also face a series of decisions. However, the decisions of **PLAYER B** are hypothetical and will not affect his/her own earnings nor your earnings. **Your decisions** will **entirely determine your own earnings and PLAYER B's earnings**.

Selection of decision situation relevant for payout

- At the end of the experiment **one** of the decision situations in PART 1 will be randomly selected by the computer. Each decision situation is equally likely to be selected. Your decision in the selected decision situation will determine your earnings and the earnings of **PLAYER B**. **Therefore, you should consider each decision situation independently and as the decision situation that counts for the earnings to YOU and to PLAYER B.**

- The bank transfers with the money that **YOU** and **PLAYER B** earn in PART 1 will be made **in 5 weeks from today**.

Your identity

- You will never get to know the identity of the participants in the role of **PLAYER B** nor will these participants ever get to know your identity. Your anonymity is thus guaranteed.

Short summary

- **YOU** are assigned to the role of **PLAYER A** throughout the whole experiment.
- At the end of the experiment, one of the decision situations in PART 1 will be randomly selected to be the one relevant for the earnings that **YOU** and **PLAYER B** will receive. Each decision situation is equally likely to be selected. Therefore you should treat each decision situation as the one that counts for earnings.
- **YOU** will face several decision situations where you have to decide how to allocate an amount of tokens between **YOU** and **PLAYER B**.
- In each decision situation, **YOU** will be randomly rematched with one participant who is in the role of **PLAYER B**. It is very unlikely that you will be matched with the same **PLAYER B** more than once.
- The **amount of tokens** you are asked to divide could **vary** across decision situations.
- The **value** of each tokens you decide to allocate to **YOU** and to **PLAYER B** could **vary** across decision situations.
- At the end of the experiment, the points **YOU** and **PLAYER B** earn are exchanged to Euro at the rate of 1 point = 0.10 €.
- Earnings will be transferred to your and **PLAYER B**'s bank accounts. Earnings will be transferred exactly in 5 weeks.
- Your decisions **will entirely determine your earnings and the earnings of PLAYER B**.

If you have any questions please do not ask them aloud but raise your hand. One of the experimenters will come to you to answer your question in private.
If you have no questions and you are ready to start with the experiment, please click Continue on the screen and you will be asked few comprehension questions.

7.5. *Instructions Intertemporal Task*

Welcome to PART 2 of the experiment. **In this part of the experiment the decisions you make will only affect YOUR earnings.**[30]

In this part, you are asked to make a series of **36** decisions about how to divide **100 tokens** between two dates, an earlier date and a later date. Tokens will be exchanged for Euro. The exchange rate varies between the earlier date and the later date, and varies from one decision situation to the next. The tokens you allocate to the later date will never be worth less money than tokens you allocate to the earlier date. The difference in value of tokens will depend on the decision situation. The decision situations will be displayed in six **Decision Forms**. Each Decision Form contains six decision situations. Each decision situation is represented as a row in a Decision Form. **Figure 3.9** below represents an example of a decision situation you may face during this part of the experiment. In the example below **'today' is the 9th of August 2017**. In the experiment 'today' will refer to the actual date of today.

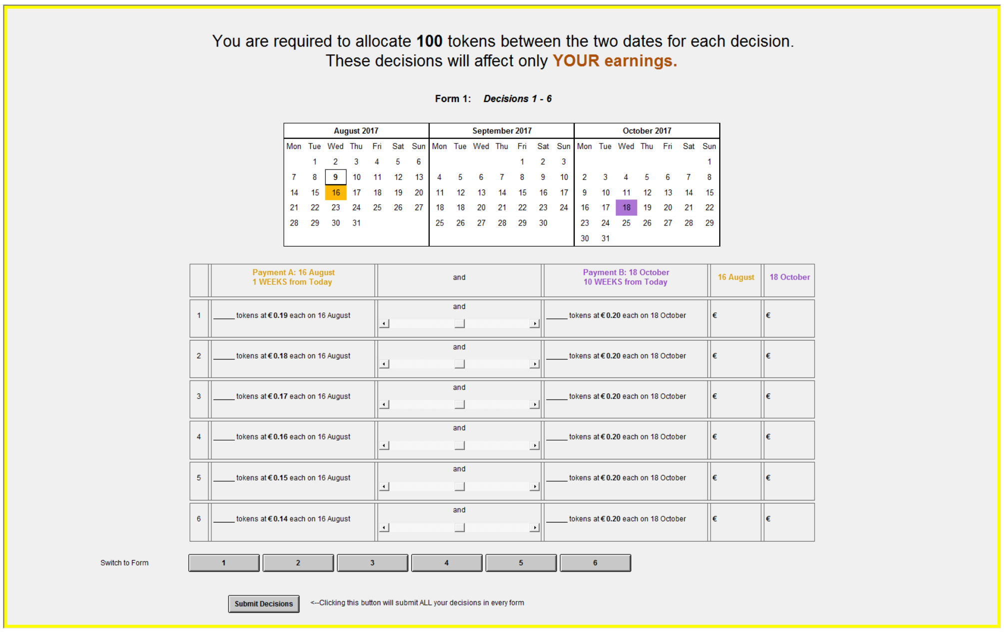


Figure 3.9: Decision Situation - Example 1

In this example, the first decision situation on the screen shows the choice to allocate **100 tokens** between **16 August** and **18 October**.

- On **the top of the screen** you see a calendar.
  Note that today's date (in this example the 9 August) is always highlighted in **bold** and is surrounded by a square shape. Also note that the **earlier date** (**16 August**) is highlighted in **Yellow** while the **later date** (**18 October**) is highlighted in **Purple**. In each decision situation the dates are highlighted so that you can easily see what the earlier date and later date are.

- In **the lower part** of the screen you see a table.
  The column in the middle of the table displays a slider you should use to allocate tokens between the earlier date and the later date. To make your allocation decision, you have to either scroll the bar or click the arrows. The computer will automatically calculate the earnings you will receive on both the earlier and the later dates. The money you earn at the earlier date and at the later date are displayed in the last two columns on the right of the table.

[30]In the experiment, the order of the intertemporal task for *Self* and for *Other* was counterbalanced.

Remember there are **six** Decision Forms in this part. You can switch between Decision Forms by pressing the button with the number of the Decision Form. You can revise your choices as often as you like, within and across Decision Forms. Once you are satisfied with all of your choices, you can click on the "Submit Decisions" button to submit ALL your decisions in the six Decision Forms.

In this example, if you consider the first decision situation, each token you allocate to **16 August** is worth 19 cents, while each token you allocate to **18 October** is worth 20 cents.
Thus, if you allocate 100 tokens to **16 August** you will earn 100 x 0.19€ = 19€ on this date, and nothing on **18 October**.
If you allocate all 100 tokens to **18 October** you will receive 100 x 0.20€ = 20€ on this date and nothing on **16 August**.
You are also free to allocate some tokens to earlier date and some to the later date. For instance, if you allocate 64 tokens to **16 August** and 36 to **18 October**, you will earn
64 x 0.19€ = 12.16€ on **16 August** and 36 x 0.20€ = 7.2€ on **18 October**.

Note: *In different decision situations, the exchange rate of tokens allocated to the earlier date will vary.*

Determination of Earnings - PART 2

- At the end of the experiment **one** of the decision situations in PART 2 will be randomly selected by the computer. Each decision situation is equally likely to be selected. Your decision in the selected decision situation will determine your earnings in this part of the experiment. **Therefore you should treat each decision situation as the one that counts for YOUR earnings.**
- In addition to the earnings from the selected decision, **YOU** will always receive 1€ on the earlier date and 1€ on the later date. Hence, **YOU** will always receive payments twice, even if you allocate 0 tokens to one of the two dates.
- The bank transfers with the money that **YOU** will earn in PART 2 will be made exactly on the scheduled days as in the decisions situations.
- The decisions in this part of the experiment affect only **YOUR earnings**.

Short summary

PART 2:

- You will face **36** decision situations where you have to decide how to allocate an amount of tokens between an earlier date and a later date.
- The decisions in this part of the experiment affect only **YOUR earnings**.
- The **exchange rate of tokens** to Euro could **vary** across decision situations.
- The **days** of the earlier and later dates could **vary** across decision situations.
- At the end of the experiment, **one** of the decision situations in PART 2 will be randomly selected to be the one relevant for determining your earnings. Each decision situation is equally likely to be selected. **Therefore you should treat each decision situation as the one that counts for your earnings.**
- Earnings will be transferred to your bank account. Earning will be transferred exactly on the scheduled days as in the decisions.
- In addition to the earnings from the selected decision, **YOU** will always receive 1€ on the earlier date and 1€ on the later date. Hence, **YOU** will always receive payments twice, even if you allocate 0 tokens to one of the two dates.

Detailed instructions - PART 3

Welcome to PART 3 of the experiment. This part is similar to PART 2. However, this time you will make a series of **36** decisions for another participant, a randomly chosen **PLAYER B. In this part of the experiment the decisions you make will affect only the earnings of the selected PLAYER B.**
As in PART 2, in this part you are asked to make a series of **36** decisions about how to divide 100 tokens between two dates, an earlier date and a later date. Tokens will be exchanged for Euro. The exchange rate varies between the earlier date and the later date, and varies from one decision situation to the next. The tokens you allocate to the later date will never be worth less money than tokens you allocate to the earlier date. The difference in value of tokens will depend on the decision situation. The decision situations will be displayed in six **Decision Forms**. Each Decision Form contains six decision situations. Each decision situation is represented as a row in a Decision Form.
The difference from PART 2 is that now you are asked to make decision for a randomly selected PLAYER B.
Figure 3.10 below represents an example of a decision situation you may face during this part of the experiment. In the example below **'today' is the 9th of August 2017**. In the experiment 'today' will refer to the actual date of today.

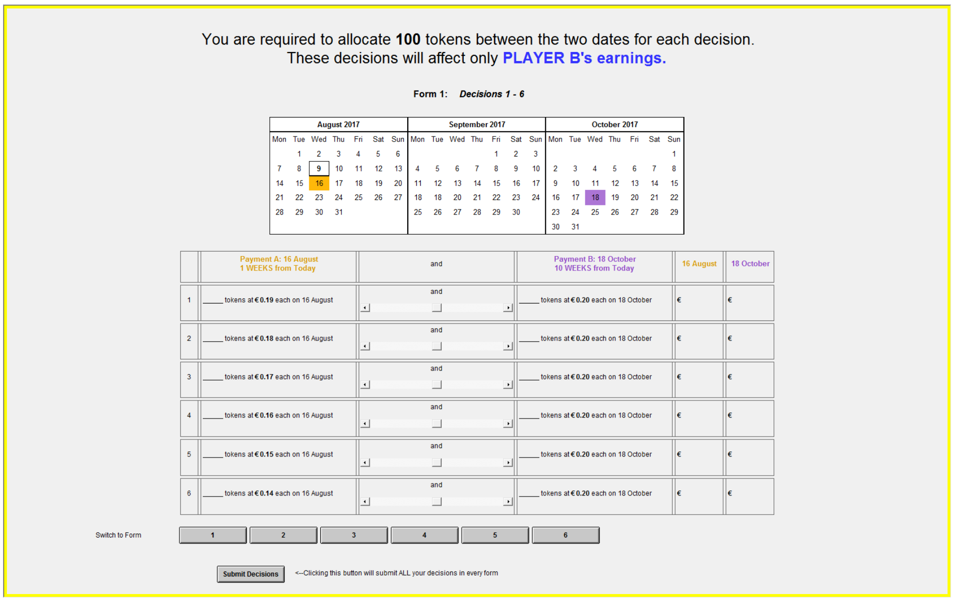


Figure 3.10: Decision Situation - Example 1

In this example, the first decision on the screen shows the choice to allocate **100 tokens** between **16 August** and **18 October**.

- On **the top of the screen** you see a calendar.
  Note that today's date (in this example 9 August) is always highlighted in **bold** and is surrounded by a square shape. Also note that the **earlier date** (**16 August**) is highlighted in **Yellow** while the **later date** (**18 October**) is highlighted in **Purple**. In each decision situation the dates are highlighted so that you can easily see what the earlier date and later date are.

- In **the lower part** of the screen you see a table.
  The column in the middle of the table displays a slider you should use to allocate tokens between the earlier date and the later date. To make your allocation decision, you have to either scroll the bar or click the arrows. The computer will automatically calculate the earnings **PLAYER B** will receive on both the earlier and the later dates. The money **PLAYER B** earns at the earlier and at later dates

are displayed in the last two columns on the right of the table.
Remember that in this part **you are making decisions for a randomly selected PLAYER B**. The money he/she earns in the earlier date and in later date are displayed in the last two columns on the right of the table.
There are six Decision Forms in this part. You can switch between Decision Forms by pressing the button with the number of the Decision Form. You can revise your choices as often as you like, within and across Decision Forms. Once you are satisfied with all of your choices, you can click on the "Submit Decisions" button to submit ALL your decisions in the six Decision Forms.

In this example, if you consider the first decision situation, each token you allocate to **16 August** is worth 19 cents, while each token you allocate to **18 October** is worth 20 cents.
Thus, if you allocate 100 tokens to **16 August**, **PLAYER B** will earn 100 x 0.19€ = 19€ on this date, and nothing on **18 October**.
If you allocate all 100 tokens to **18 October**, **PLAYER B** will receive 100 x 0.20€ = 20€ on this date and nothing on **16 August**.
You are also free to allocate some tokens to earlier date and some to the later date. For instance, if you allocate 64 tokens to **16 August** and 36 to **18 October**, **PLAYER B** will earn
64 x 0.19€ = 12.16€ on **16 August** and 36 x 0.20€ = 7.2€ on **18 October**.

Note: *In different decision situations, the exchange rate of tokens allocated to the earlier date will vary.*

Determination of earnings - PART 3

- You will make decisions for another participant, a randomly selected **PLAYER B**.
- At the end of the experiment **one** of the decision situations in PART 3 will be randomly selected by the computer. Each decision situation is equally likely to be selected. Your decision in the selected decision situation will determine the earnings of **PLAYER B**. **Therefore you should treat each decision situation as the one that counts for his/her earnings.**
- In addition to the earnings from the selected decision, **PLAYER B** will always receive 1€ on the earlier date and 1€ on the later date. Hence, he/she will always receive payments twice, even if you allocate 0 token to one of the two dates. Except from these 2€, your decision in the selected situation will entirely determine the earnings of **PLAYER B**.
- The earnings of **PLAYER B** in PART 3 will be transferred exactly on the scheduled days as in the decisions.

Your identity

- You will never get to know the identity of the participants in the role of **PLAYER B** nor will these participants ever get to know your identity. Your anonymity is thus guaranteed.

Short summary

PART 3:

- You will face **36** decision situations where you have to decide how to allocate an amount of tokens between two dates, an earlier date and a later date for another participant assigned to the role of **PLAYER B**.
- The decision in this part of the experiment affect only the earnings of the selected **PLAYER B**.

- The **exchange rate of tokens** to Euro could **vary** across decision situations.
- The **days** of the earlier and later dates could **vary** across decision situations.
- At the end of the experiment, **one** of the decision situations in this part will be randomly selected to be the one relevant for determining **PLAYER B**'s earnings. Each decision situation is equally likely to be selected. Therefore you should treat each decision situation as the one that counts for **PLAYER B**'s earnings.
- In addition to the earnings from the selected decision, **PLAYER B** will always receive 1 € on the earlier date and 1 € on the later date. Hence, he/she will always receive payments twice, even if you allocate 0 token to one of the two dates. Except from these 2 €, your decision in the selected situation will entirely determine the earnings of **PLAYER B**.
- The earnings will be transferred to **PLAYER B**'s bank account. The earning will be transferred exactly on the scheduled days as in the decisions.

Chapter 4

In search of the sunk cost bias

Abstract

Evidence from hypothetical scenarios strongly suggests the existence of a sunk cost bias, the tendency to 'throw good money after bad money'. However, the few studies using incentives are inconclusive. In addition, evidence on potential psychological channels underlying such a bias is scarce. We present a laboratory experiment designed to investigate the sunk cost bias and to test some prominent psychological mechanisms. Inspired by the hypothetical scenarios, we use a two-stage investment task in which an initial investment needs to be made to start a project. In the initial investment stage, the size of the investment and the responsibility of the investor are exogenously varied. In the second investment stage, participants can either decide to terminate the project or to make an additional investment to finish the project. We do not find evidence for the sunk cost bias. To the contrary, we observe a robust *reverse* sunk cost bias. That is, the larger the initial investment, the lower the likelihood to continue investing in a project. Moreover, whether or not subjects are responsible for the initial investment, does not affect their additional investment. More waste aversion individuals also do not react more strongly to sunk cost whereas being in the loss domain decreases additional investment. Importantly, we replicate the sunk cost bias when using hypothetical scenarios. However, the reverse sunk cost bias also holds for those participants who exhibit a strong sunk cost bias in the hypothetical scenarios.

This paper is co-authored with Matthias Wibral & Arno Riedl.

1. Introduction

The sunk cost bias refers to the behavioral tendency to continue an endeavor once an investment has been made, even if it is not optimal to do so (Arkes and Blumer, 1985). The bulk of the evidence suggesting the existence of the sunk cost bias consists of responses to hypothetical survey questions (e.g. Arkes and Ayton, 1999; Fox and Staw, 1979; Molden and Hui, 2011; Soman and Cheema, 2001; Staw, 1976; Strough et al., 2008). For instance, in the seminal paper by Arkes and Blumer (1985), participants are asked to imagine that they are the owners of a company and have previously invested a large sum of money in what seemed to be a promising project. When the project is almost finished, they learn that a competitor is about to release a better product at a cheaper price. Respondents then need to consider whether to stop investing in the development of their product and realize the loss or to persist with the project by making an additional investment. Participants overwhelmingly state to carry out the additional investment and are thus considered to fall prey to the sunk cost bias.

Examining the sunk cost bias is important because it has been implicated in a wide spectrum of situations involving sunk costs in practice. For example, the sunk cost bias has been put forward as an explanation for why politicians continue public works that went over budget (Ross and Staw, 1993), why firms continue to invest in hopeless projects (Arkes and Blumer, 1985), why people stay in failing relationships (Strube, 1988), or why researchers continue less promising projects instead of starting new ones.

Interestingly, despite the intuitive appeal of the concept and the substantial body of evidence from survey studies, it has been hard to demonstrate the sunk cost bias in incentivized studies, both in the field and the lab (e.g. Friedman et al., 2007; Ashraf et al., 2010). Existing attempts to study the sunk cost bias in the laboratory often have quite complicated designs and may be prone to game form misconceptions (Cason and Plott, 2014), which might explain the inconsistent evidence. Therefore, it is important to have a workhorse for studying the bias that is simple to understand and easily implementable, yet rich enough to allow learning more about the psychological mechanisms underlying the sunk cost bias.

In this paper we present a novel design with incentivized choices to investigate the sunk cost bias as well as important potential psychological mechanisms that could drive the bias.[1] The experimental design is inspired by the classic project continuation example from the survey literature. Specifically, we study a two-stage investment task in which an initial investment needs to be made to start a project and to advance it to a second investment stage. In the second investment stage, participants know the size of the initial investment size and can either decide to terminate the project, or to carry out an additional investment to finish the project. If an additional investment is made, the project is successful and yields a high payoff with some known probability.

When participants make the decision about the initial investment, the exact cost of the initial investment are unknown to them, but they do know the distribution of the potential costs. Participants learn the exact costs of their initial investment only when they have to decide whether to make the additional investment. The key idea here is that by varying the amount

[1]The study reported in the paper was preregistered at OSF (https://osf.io/c253e).

initially invested (i.e., the sunk cost), we can study the impact of the size of the sunk cost on the willingness to make the additional investment. A rational decision-maker should ignore the initial investment and decide whether to make the additional investment based solely on the expected utility of the available options. If we observe that participants are more willing to make the additional investment after larger initial investments, this is evidence in favor of the sunk cost bias.

Our design also includes features that allow us to test several psychological mechanisms that have been proposed as drivers of the sunk cost bias. Specifically, we examine the roles of responsibility for the initial investment, of waste aversion, and of being in the loss domain when making the additional investment decision. Concerning the role of responsibility, self-justification and cognitive dissonance theory (e.g. Bazerman et al., 1984; Brockner, 1992; Staw, 1976) propose that abandoning a project after an initial investment requires admitting that the initial investment was a bad decision. The sunk cost bias arises because continuing to pour resources into a failing course of action is a way to justify one's own past decisions. Personal responsibility for the sunk cost should thus increase the willingness to invest additional resources for the continuation of a project. To test this, we compare two types of situations, one in which participants are responsible for the initial investment, and one in which they are not.

Another reason why the sunk cost bias may occur could be waste aversion. Arkes and Blumer (1985) suggest that people are more willing to invest after bad news because not investing constitutes an admission that the prior expenses were wasted. To investigate this, we include a questionnaire measure of waste aversion (Haller and Schwabe, 2014). If waste aversion drives the sunk cost bias, we would expect that more waste averse participants display a stronger sunk cost bias independent of responsibility for the initial investment. In addition, the effect should increase with the size of the sunk cost.

Our exogenous variation of sunk cost at the individual level also allows us to test whether the sunk cost bias depends on being in the loss domain. Prospect Theory suggests that the value function is concave in the gain domain and convex in the loss domain, relatively to a reference point (Kahneman and Tversky, 1979). According to this S-shaped value function, individuals are risk seeking in the domain of losses and risk averse in the gain domain. Being in the loss domain may thus lead to the sunk cost bias because further losses do not result in large decreases in value; however, comparable gains do result in large increases in value.

There are two additional features of our design. First, it is simple and easy to understand. As we discuss below, some of the previous work in experimental economics has used relatively complicated designs which may have confused subjects (Weigel, 2018). Second, we also replicate classic survey measures of the sunk cost bias from the psychology literature (Arkes and Blumer, 1985) in a post-experimental questionnaire. We can thus compare the results from our incentivized task to classic survey measures within subject. This comparison can shed light on the discrepancy of findings between surveys and incentivized studies.

Most of the evidence in favor of the sunk cost bias comes from hypothetical scenarios, whereas the evidence from incentivized studies is mixed (for reviews, see, e.g., Sleesman et al., 2012; Roth et al., 2015). To account for the inconsistent evidence, alternative explanations have been suggested. For instance, responses in hypothetical vignette scenarios that have been interpreted as a sunk cost bias may actually be due to the fact that participants do not fully adopt the preferences described in the scenarios, but use their own homegrown preferences (Friedman et al., 2007).[2] In addition, as noted by Weigel (2018), subjects who are indifferent or confused might exhibit choices that could be misinterpreted as the sunk cost bias.

Regarding field data, three recent papers report evidence that is consistent with a sunk cost bias. Augenblick (2016) shows that data from penny auctions are in line with the predictions of a theoretical model in which players' value of winning the good increases with their previous bid costs. Ratnadiwakara and Yerramilli (2017) find that past property taxes in California lead to a significant increase of the sellers' chosen listing price. Ho et al. (2018) exploit changes in the price of a government license to buy a car in Singapore and find that an increase in sunk costs (i.e., the price of the license) leads to an increase in driving.

Attempts to find the sunk cost bias in field experiments have been less successful. Arkes and Blumer (1985) find that randomly providing discounts to buyers of theatre season tickets decreases show attendance. However, this effect is only observed for the first half of the theatre season and the sample is quite small. Ashraf et al. (2010) conducted field experiments testing the impact of transaction price on the usage of a certain product. They report no evidence of a sunk cost bias as households paying a higher transaction price are not more likely to use the product. Ketel et al. (2016) test for a sunk cost bias in an educational setting. Students signing up for extra-curricular tutorial sessions randomly received a discount on the tuition fee. The authors find that on average the discount does not affect attendance or performance.

Overall, the picture emerging from field data and field experiments is thus quite mixed. One potential reason is that it is hard to fully control for the possible confounds of selection, reputation, and subjective beliefs (Weigel, 2018). Mcafee et al. (2010) also argue that conditioning behavior on sunk costs could be rationalized, if agents react to sunk costs because of informational content, reputation, or financial and time constraints.

There are only a few incentivized laboratory experiments investigating the sunk cost bias. To our knowledge, the first attempt in economics is by Phillips et al. (1991). They study whether identical lottery tickets are valued differently depending on the price (i.e., the sunk cost) at which they were bought. Only a quarter of participants value the ticket more as the price increases and thus exhibit a sunk cost bias, while another quarter of subjects show a reverse sunk cost bias, as they value the ticket less with an increased price. In additional treatments with opportunities to learn in a market environment, very few subjects exhibit a sunk cost bias. Heath (1995) studies investments in a lottery where subjects can invest again in the same lottery when they have lost. He finds a reverse sunk cost bias when the sum of incremental and sunk cost leads to an overall loss. Offerman and Potters (2006) show that higher entry fees paid to gain the right to operate in a market lead some players to set prices in a more collusive way.

The most comprehensive laboratory investigation of the sunk cost bias to

[2]For example, a common scenario tells participants that they have accidentally booked vacations for the same date at two different locations and now have to decide where to go. They are told to imagine that they spent more money on location A, but that they actually prefer location B. Choosing location A is interpreted as evidence for the sunk cost bias. However, if the participant actually prefers A over B and chooses A for this reason then the response is not necesarily biased.

date is Friedman et al. (2007). The authors use a computer game where participants decide whether to keep digging for a treasure on a virtual island or to incur a cost to move to another island. They only find a very small sunk cost bias that is inconsistent across treatments with some treatments even showing a reverse sunk cost bias. Our design is similar in spirit to two recent neuroeconomics papers (Bogdanov et al., 2017; Haller and Schwabe, 2014). However, these papers use deception in a way that makes their findings hard to interpret. In their setup, both a rational decision-maker and an individual prone to the sunk cost bias may exhibit the same behavior.[3] Haita-Falah (2017) studies a potential sunk cost bias in a setup in which the channels of loss aversion and cognitive dissonance cannot drive the effect. She finds weak evidence for a sunk cost bias, which is significant only for high sunk costs. Similar to Heath (1995), Weigel (2018) studies an individual decision-making task inspired by penny auctions in which subjects endogenously accumulate sunk cost. In contrast to Heath (1995), he reports that sunk costs increase the decision maker's willingness to continue along an unprofitable course of action. Finally, Ronayne et al. (2020) find that 23% of participants in an MTurk study who have worked for a ticket for a certain lottery do not switch to a ticket for a dominant lottery when offered to do so.

Our experiment differs from all previous studies in at least one of the following dimensions. First, our design allows us to test the impact of different levels of sunk cost within subjects. Second, it considers both situations where the participant is in the gain domain and situations in which the participant is in the loss domain, relative to the initial endowment. Third, we use exogenous variation to test the role of responsibility in a clean way at the individual level. We have a larger sample per treatment and a simpler design than previous studies.[4] Fourth, we are also able to study the role of waste aversion. Finally, our design differs from previous experimental studies (except Ashraf 2010 and Ketel et al., 2016) as it allows us to compare the tendency to exhibit the sunk cost bias in both an incentivized and a hypothetical setting.

Our results can be summarized as follows. First, in contrast to studies using hypothetical questions and all other incentivized studies (except for Heath, 1995), and in contrast to our pre-registered hypotheses, we do not observe a sunk cost bias but a *reverse* sunk cost bias. That is, the larger the initial investment, the lower the likelihood of continuing to invest in a project. This result also holds when we only consider those participants who display the sunk cost effect in the hypothetical choice scenarios. Second, contrary to our hypothesis, we find no difference in behavior depending on whether subjects are responsible for the initial investment or not. That is, the reverse sunk cost bias is observed irrespective of whether or not participants are responsible for the initial investment. Third, contrary to our expectation, we observe that

[3]Haller and Schwabe (2014) and Bogdanov et al. (2017) study a setup in which a risky investment project is sometimes successful immediately after a first investment, and sometimes further investments are needed. A higher willingness to make the second investment compared to the first one is taken as evidence for a sunk cost bias. However, the true success probabilities and the stated ones differ in a way such that first investments have a lower success probability. If subjects learn this over time, then they will be less likely to make the first investment compared to the second one even if they do not exhibit a sunk cost bias.

[4]For example, Friedman et al. (2007) state: "it took us several tries over a period of months to get it [the equilibrium strategy in their experiment] right." In a replication attempt of Haita-Falah (2017) reported by Weigel (2018), numerous subjects complained that they did not understand the instructions.

participants are even less willing to continue investing when they find themselves in the loss domain compared to the gain domain. Fourth, we observe that higher self-reported waste aversion is positively correlated with the willingness to make the additional investment. However, this effect is not larger for higher higher levels of sunk cost. Finally, we replicate the findings from the hypothetical scenarios of Arkes and Blumer (1985). Participants generally exhibit the sunk cost bias in hypothetical scenarios, and especially when they imagine they were responsible for them. Our reverse sunk cost bias in the incentivized settings thus does not seem to be due to an idiosyncratic sample. We do not find support that the sunk cost bias in the hypothetical scenarios translates into a sunk cost bias in the incentivized investment task. Even those participants who exhibit a strong sunk cost bias in the hypothetical scenarios show a reverse sunk cost bias in the incentivized investment task.

The rest of the paper is organized as follows. Section 2 presents the experimental design. Section 3 describes the pre-registered testable hypotheses. Section 4 shows the experimental results. Finally, Section 5 discusses the results and limitations of the study and concludes.

2. Experimental design

The main part of the experiment consisted of a repeated choice task comprising two investment stages, which was followed by a questionnaire and an additional choice task. Instructions were given prior to the start of each respective part. The questionnaire included questions on sociodemographic characteristics, waste aversion, and hypothetical scenarios (adapted from Arkes and Blumer, 1985) to elicit the sunk cost bias. After the questionnaire, participants played a lottery choice task developed by Gächter et al. (2007) to measure individual loss aversion. The complete set of instructions used in the experiment (including screenshots) are provided in the Appendix.

2.1. Investment task

The investment task is framed as investing in a project and is inspired by project continuation scenarios used in some of the questionnaire studies reporting a sunk cost bias. It consists of two investment stages: an *initial investment stage* and a *continuation stage*. Figure 4.1 provides an overview of the setup of the investment task. In total, each participant makes decisions in 36 investment tasks, which are evenly split into so-called *Responsibility* projects and *No Responsibility* projects. These projects are presented in random order and participants do not receive any feedback regarding the success of the project during the task. In the following, we describe the set-up of a *Responsibility* project as depicted in Figure 4.1.

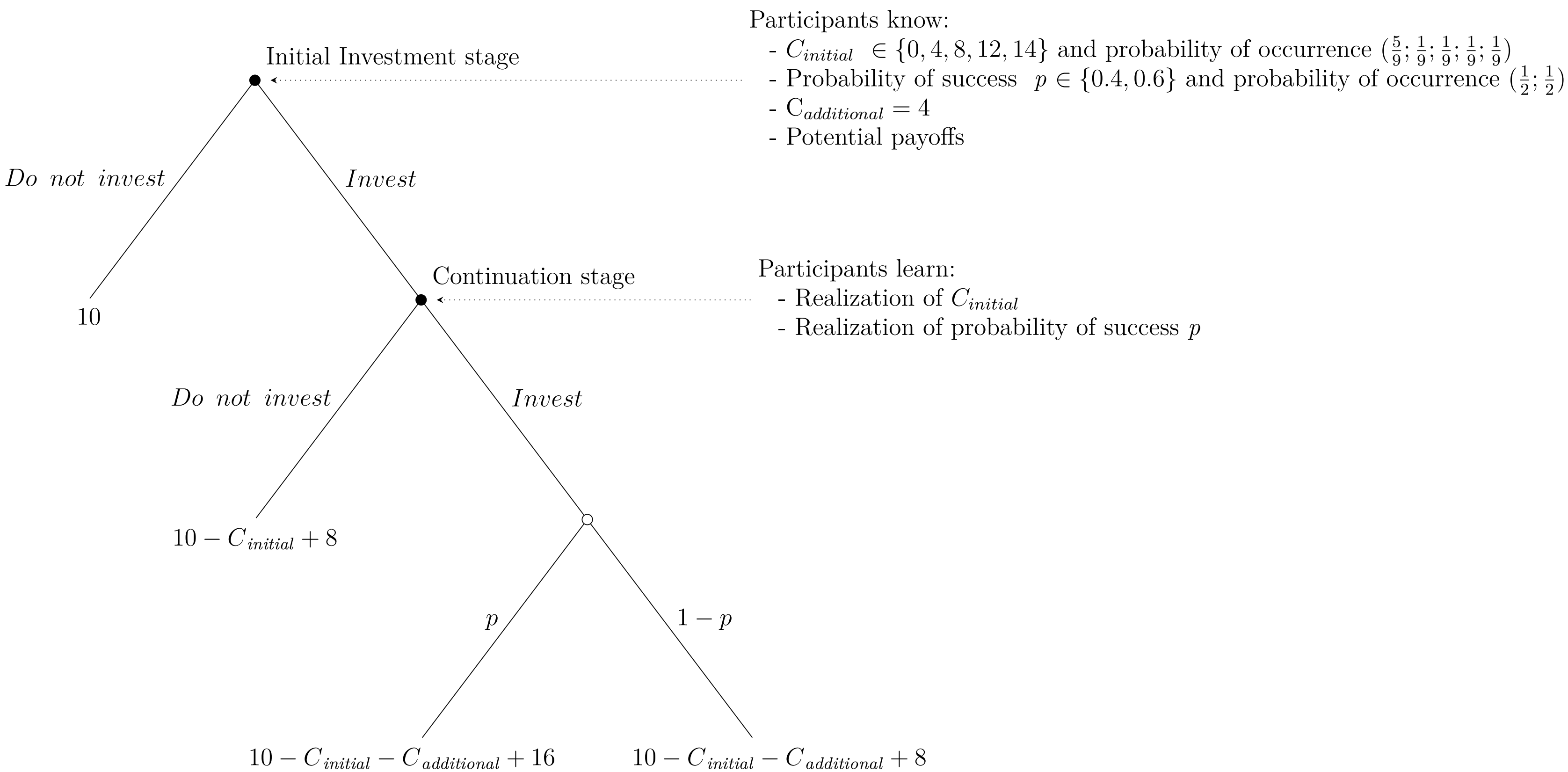


Figure 4.1: **Decision tree of an investment task** (*Responsibility* project)

Responsibility projects. In the *initial investment stage*, participants receive an endowment of €10 and decide whether to start a project of which they know that it has a positive expected value.[5] Participants know the distribution of the potential costs $C_{initial}$ of the initial investment, but the realization of the costs is ex-ante unknown to them. $C_{initial}$ is €0 in 10 out of the 18 projects (*No Initial Costs Projects*), and €4, €8, €12 and €14, respectively, in two out of the 18 projects (*Initial Costs Projects*).[6] At this stage participants also know that if they start a project they will move on to the *continuation stage* where they can decide to continue the project or not. They are also informed that the investment cost to continue the project will be $C_{additional}$ = €4 and that there are two success probabilities, $p = 0.4$ or $p = 0.6$, which each will occur with 50% chance. In addition, participants know all potential outcomes and, thus, earnings.

In the initial investment stage, those participants who decide not to make the initial investment earn the initial endowment of €10 and move to the initial investment stage of the next investment task. Those who decide to make the initial investment learn about the realization of the associated $C_{initial}$ and move to the continuation stage. In the *continuation stage*, they are reminded of the initial investment costs $C_{initial}$ that are payed already and that the additional investment is associated with a known cost $C_{additional}$ = €4. They also learn about the actual success probability of the project.[7] Participants also know all potential earnings associated with their decisions. Given this information, participants have to decide whether to continue the project by making an additional investment, or to abandon it.

The different payoffs associated with different courses of action in the continuation stage are as follows. Participants receive €8 if they decide not to make the additional investment. Thus, a participant who makes the initial investment, but then abandons the project in the continuation stage earns €10 - $C_{initial}$ + €8 = €18 - $C_{initial}$. If participants make the additional investment and the project is unsuccessful, the investment returns €8. In case the project is successful, the investment returns €16. Thus, a participant who makes the additional investment in the continuation stage earns €10 - $C_{initial}$ – €4 + €8 = €14 - $C_{initial}$ if the project is unsuccessful, and €10 - $C_{initial}$ – €4 + €16 = €22 - $C_{initial}$ if the project is successful.

No Responsibility projects. In the *No Responsibility* projects, participants are also initially endowed with €10, but here the computer makes the initial investment on their behalf. Thus, in this type of projects, participants are not responsible for the initial investment and the associated sunk cost. After the initial investment is made, participants are presented with the continuation stage, which is identical in both the *Responsibility* and the *No Responsibility* projects.

[5]Participants know the decision tree and thus can calculate the expected value of an initial investment and later also of the additional investment.

[6]Participants did not know the exact number of decision situations for each level of sunk cost. They were informed that: "If you make the initial investment, there will be an investment cost associated to your decision. You will know the actual investment cost only after you have made the initial investment. On average, you can expect the initial investment cost to be: €0 five out of nine times; €4 one out of nine times; €8 one out of nine times; €12 one out of nine times; €14 one out of nine times".

[7]In fact, each success probability occurs in half of the investment tasks in a random order across decision situations for each individual.

We did not use the labels *Responsibility* and *No Responsibility* in the experiment. Participants were informed that in some cases they could decide to make the initial investment or not, and in other cases the computer would make the initial investment on their behalf. In the experiment, we implemented the decision screen such that at the beginning of the initial investment stage in the *Responsibility* projects participants were reminded of the distribution of $C_{initial}$ and were given the opportunity to make the initial investment or not. In the *No Responsibility* projects, participants were again reminded about the distribution of $C_{initial}$, and informed that the computer made the decision on their behalf. The decision screens can be found in the Appendix.

Several considerations guided our choice of parameter values. First, as we are interested in the behavior in the continuation stage, we chose the parameters such that making the initial investment is attractive in expectation for moderately risk averse participants. The expected value of making the initial investment is €14.18 and thus substantially higher than the endowment of €10. Ex-ante only someone with very strong risk aversion should not make the initial investment.[8] We chose the distribution of $C_{initial}$ (i.e., the sunk cost) such that the probability of $C_{initial}$ = 0 is above 50% with the goal to make the initial investment decision attractive, while at the same time still having a fairly equal number of observations with $C_{initial}$ > 0.

Second, we chose the different levels of the sunk cost such that for half of the decision situations in which there is a strictly positive sunk cost, abandoning the project in the continuation stage does not lead to a loss compared to the endowment. For $C_{initial}$ of €4, participants who decide to abandon the project in the continuation stage receive a payoff of €14 and are better off than if they had not made the initial investment and received their endowment of €10. For a $C_{initial}$ of €8, they break exactly even. For the other two levels of $C_{initial}$ (€12 and €14), participants make a loss compared to their endowment if they decide to abandon the project. These values were chosen such that assuming a reference point of €10, given the median estimated curvature parameters, loss aversion coefficient, and probability weighting values from Tversky and Kahneman (1992), Cumulative Prospect Theory would predict the following: a participant will make the additional investment with success probability $p \geq 0.4$ only for $C_{initial} \geq$ €12.[9] Third, we chose to have two different success probabilities in the continuation stage to minimize possible floor and ceiling effects, thus having a greater chance of finding a sunk cost bias.

---

[8]For example, any decision maker with the CRRA utility function $U(x) = \frac{1}{1-r}x^{1-r}$ with an r < 2.05 will always make the initial investment. For comparison, Holt and Laury (2002) do not observe any participant with r > 1.37.

[9] We apply Cumulative Prospect Theory as in Tversky and Kahneman (1992), where the prospect is the product of decision weights $\pi(p)$ and value of the potential outcome, as shown by: $V(x) = \begin{cases} x^{\alpha}\ if\ x \geq 0 \\ -\lambda(-x)^{\beta}\ if\ x < 0 \end{cases}$. We assume a reference point of €10 (i.e., the initial endowment) and the parameters as estimated in Tversky and Kahneman (1992) of loss aversion $\lambda = -2.25$, curvature coefficients in the positive domain (α) and in the negative domain $(\beta) = 0.88$, and a decision weight $\pi(p) = \frac{\rho^{\gamma}}{(\rho^{\gamma}+(1-\rho)^{\gamma})^{\frac{1}{\gamma}}}$ with $\gamma = 0.61$ in the gain domain and $\gamma = 0.69$ in the loss domain.

2.2. *Psychological measures related to the sunk cost bias*

Waste aversion has been proposed as an explanation for the sunk cost bias (Arkes and Blumer, 1985). To explore this mechanism, we ask participants fill out a short questionnaire (Haller and Schwabe, 2014) that aims to assess their desire not to waste resources after the investment task. This questionnaire consists of four statements that are answered on a scale from 1 ("I do not agree") to 11 ("I completely agree): "It is important for me not to appear wasteful", "Wasted investments hurt me", "People who know me think I am wasteful" (inversely coded), and "It annoys me if investments are not successful". The scores for the 4 items are summed up and the total score is taken as an indicator of the strength of the individual's desire not to appear wasteful. With our design, we can disentangle the role of waste aversion from cognitive dissonance as its effect should be present both with and without responsibility for incurring sunk costs, whereas cognitive dissonance should only affect behavior in the *Responsibility* projects.

To examine whether the survey measure of the sunk cost bias used in the literature correlates with our incentivized measure, participants also answer four binary hypothetical questions related to the sunk cost bias. These questions are slightly modified versions of four vignette studies in the seminal paper by Arkes and Blumer (1985). To make them more relatable to our subject pool, we changed the travel destinations and the monetary amounts used in the hypothetical scenarios. The scenarios are reported in full in the Appendix. For two scenarios, participants were instructed to imagine they were responsible for the initial investment, while for the other two scenarios they were told someone else was responsible for it. From the answers, we construct an index ranging from 0 to 4 indicating the strength of the tendency to show the sunk cost bias in hypothetical scenarios.

2.3. *Loss aversion*

To measure loss aversion, we use a task by Gächter et al. (2007).[10] In this task, participants decide for each of six lotteries whether they want to accept it or reject it. Across lotteries, the winning prize is fixed at €6 while the potential loss varies between €2 and €7, as shown in Table 4.12 in the Appendix.

Under Cumulative Prospect Theory, behavior in this task is jointly determined by probability weighting, the curvatures of the utility function in the gain and loss domain, and loss aversion. Under certain assumptions, in particular, linearity of the value function, the task provides a simple measure of the loss aversion parameter in cumulative prospect theory.[11] In the context of the sunk cost bias, the linearity assumption can influence predictions

[10]To keep the experiment within reasonable time limits, we decided against a full-blown estimation of the parameters of a prospect theoretic utility function (e.g., Sokol-Hessner et al., 2009)

[11]Gächter et al. (2007) assume that a participant is indifferent between accepting and rejecting the lottery if $w^{+}(0.5)v(G) = w^{-}(0.5)\lambda\, v(L)$, where L denotes the loss in a given lottery and G the gain; $v(x)$ is the utility of the outcome $x \in \{G, L\}$, λ denotes the coefficient of loss aversion in the choice task; and $w^{+}(0.5)$ and $w^{-}(0.5)$ denote the probability weights for the 0.50 chance of gaining G or losing L, respectively. Considering that $w^{+}(0.5) = w^{-}(0.5)$, only the ratio $\frac{v(G)}{v(L)} = \lambda$ defines an individual's implied loss aversion in the lottery choice task. The additional assumption that $v(x)$ is linear ($v(x) = x$) for small amounts yields a very simple measure of loss aversion: $\lambda = G/L$.

substantially. For example, a loss averse individual who does not weight probabilities and has a linear value function will not display a sunk cost bias in our experiment. In fact, for a sufficiently high degree of loss aversion ($\lambda > 2$), such an individual would be less likely to make the additional investment around their reference point than when all outcomes lie either in the gain or in the loss domain. In contrast, another individual with the median cumulative prospect theory parameter values estimated in Tversky and Kahneman (1992) would show an identical choice pattern in our loss aversion task as the previous individual, but display a sunk cost bias in the investment task.[12]

In keeping with the convention, we call the switching point between acceptance and rejection of the lottery "Loss aversion", but the reader should be aware that this switching point might also reflect factors other than loss aversion such as probability weighting and curvature. We also ran all analyses using the loss aversion λ coefficient as calculated by Gächter et al. (2007), i.e., assuming that the value function is linear, instead of the switching point between acceptance and rejection lottery. All results are qualitatively robust to this alternative specification.

2.4. *Procedure*

We recruited 108 participants (42 men; mean age = 21.5 years, s.d. =2.5 years) using ORSEE (Greiner, 2015). The experiment was programmed with z-Tree (Fischbacher, 2007) and conducted at the Behavioral and Experimental Economics Laboratory (BEELab) of Maastricht University. Each participant completed the experiment in a randomly assigned cubicle isolated from other participants. In total, we conducted five experimental sessions and each session lasted about 60 minutes. Each participant completed the incentivized investment task first, followed by the psychological measures and finally by the loss aversion task. At the end of the experiment, one randomly selected decision of the investment task and one randomly selected decision in the loss aversion task counted for payment. Participants were informed that any losses in the loss aversion task would be deducted from a flat fee of €7 they earned for answering the questionnaire and the hypothetical questions reported in Section 2.2. During the investment task, participants did not know that there would be other tasks or payments later in the experiment. Participants earned €13.75 on average and all earnings were paid via bank transfers, a common procedure in the Netherlands. Subjects were truthfully informed that the payments were issued by research assistants unrelated to the experiment and subsequent data analysis, and that their anonymity was thus assured.

3. Hypotheses

The hypotheses were preregistered and can be found at (https://osf.io/c253e).[13] Our main interest lies in the additional investment decisions in the continuation stage. A rational decision-maker should base her investment choice only on the costs and benefits that are expected to arise from this choice. In our case, a rational decision-maker should base

[12]Such an individual would always make the additional investment for $C_{initial} \geq 12$, but not for $C_{initial} < 12$, assuming a reference point of 10, i.e., the initial endowment (see Footnote 9).

[13]Note that the hypotheses are the same as in the preregistration document, but their order and exact formulation have been modified for reasons of clarity.

investment decisions only on the expected utility of the investment project and ignore any sunk cost. However, if a decision-maker is prone to the sunk cost bias, the decision to make the additional investment should be affected by the existence as well as level of the initial investment cost, $C_{initial}$. Accordingly, our null hypothesis is that participants' decisions in the continuation stage are independent of the investment costs incurred in the initial investment stage. Our alternative hypotheses relate to the existence of a sunk cost bias.

Hypothesis 1 SUNK COST BIAS. *(a) Participants are more likely to make the additional investment when* $C_{initial} > €0$ *than when* $C_{initial} = €0$*. (b) The higher* $C_{initial}$*, the more likely it is that the additional investment is made.*

According to the cognitive dissonance account, we expect a stronger sunk cost bias for projects where participants are responsible for the initial investment (*Responsibility* projects) than for projects where the computer decides to make the initial investment on their behalf (*No Responsibility* projects). The responsibility effect (Staw, 1976) states that when people are confronted with negative feedback about the success of a decision, they commit more resources to this course of action and persist with it for a longer time if they were responsible for making the initial decision than if someone else made this decision. Based on the findings of the psychological literature (e.g., Staw, 1976; Bazerman et al., 1984; Kirby and Davis, 1998; Schoorman and Holahan, 1996; Schulz and Cheng, 2002), we formulate our second hypothesis.

Hypothesis 2 RESPONSIBILITY EFFECT. *(a) After an initial investment with* $C_{initial} > €0$*, the likelihood of making the additional investment is larger in* Responsibility *projects than in* No Responsibility *projects. (b) The increase in the likelihood of making the additional investment as* $C_{initial}$ *increases is larger in* Responsibility *projects than in* No Responsibility *projects. (c) When* $C_{initial} = €0$*, there is no difference in the likelihood of making the additional investment between* Responsibility *projects and* No Responsibility *projects.*

Conditional on finding that participants exhibit a sunk cost bias, we formulate further hypotheses relating to potential mechanisms behind the sunk cost bias. If the desire not to waste resources or not to appear wasteful drives the sunk cost effect (Arkes and Blumer, 1985), we expect that waste aversion correlates with the sunk cost bias.

Hypothesis 3 WASTE AVERSION. *(a) The higher the initial investment cost, the more the waste aversion score positively correlates with the likelihood of making the additional investment. (b) Hypothesis 2(c) is confirmed and, after initial investment when* $C_{initial} > €0$*, there is no difference in the likelihood of making an additional investment between* Responsibility *and* No Responsibility *projects.*

The sunk cost bias may emerge when participants fall behind their initial endowment of €10. That is, when they are in the loss domain relative to this endowment. Recall that, for projects with $C_{initial}$ of €12 and €14, participants who do not make an additional investment fall behind the initial endowment for sure, because in that case they only earn €6 and €4, respectively. Thus, based on Cumulative Prospect Theory (see Footnote 9) we formulate the following hypothesis.

Hypothesis 4 LOSS DOMAIN *(a) The likelihood of making the additional investment when* $C_{initial}$ *is €12 or €14 is higher than for lower non-zero initial investment costs. (b) This difference in the likelihood of making the additional* *investment is positively related to the loss aversion score.*

4. Results

In this section, we describe our findings in relation to the preregistered hypotheses. When a hypothesis is rejected, we present additional exploratory analyses. First, we investigate whether we find evidence of a sunk cost bias. Second, we describe the influence of responsibility for the sunk costs on the willingness to continue to invest. Third, we investigate the impact of waste aversion, and of being in the loss domain. Finally, we show whether the tendency to exhibit a sunk cost in a hypothetical setting is correlated with the findings in the incentivized setting.

We present results using both non-parametric tests and regression analyses. All reported tests are two-sided. In the regression analyses, we use logit models with standard errors clustered at the participant level. The dependent variable is the decision to make the additional investment in the continuation stage, coded as a binary variable taking on value 1 when the additional investment was made and value 0 when it was not made. Depending on the hypotheses considered, the sunk costs are coded in three different ways: (1) as a dummy variable that takes on value 1 when the initial investment cost is strictly positive (called *Initial Cost > 0*), (2) as a continuous variable with the values of $C_{initial}$ (called *Initial Cost*), and (3) as a dummy variable that takes on value 1 if $C_{initial}$ is €12 or €14 (called *Loss Domain*).

In the presentation of regression results we focus on those specifications where control variables are included, but also report the results without control variables for completeness. Control variables comprise the probability of project success, the measure of loss aversion and waste aversion, the score in the hypothetical sunk cost questionnaire, gender, whether the field of study is economics, and the repetitions of trials in the task, unless otherwise specified.[14] The coefficient estimates of the control variable estimates are reported in the Appendix.

4.1. Is there evidence of a sunk cost bias?

According to the sunk cost hypothesis, we expected that participants should be more likely to make the additional investment with $C_{initial} > €\,0$ (Hypothesis 1a). We also expected that the higher $C_{initial}$, the more likely a participant is to make the additional investment (Hypothesis 1b). In contrast, however, we observe that participants are less willing to make the additional investment when $C_{initial} > €\,0$ (61%) compared to when $C_{initial} = €\,0$ (79%). This difference in the relative frequency of investment choices in the continuation stage is statistically significant ($N = 108$, Wilcoxon signed-rank test, $p < 0.001$). This result is corroborated by the negative and significant coefficient of the dummy variable $C_{initial} > €\,0$ in Model (1) of the regression analysis in Table 4.1 ($p < 0.001$). Thus, our result shows a pattern opposite to the hypothesis, consistent with a *reverse* sunk cost bias.

[14]To measure loss aversion, we use the switching point in the loss aversion task described in Section 2.3. We code the variable such that a lower value implies lower loss aversion. Using the switching point in the loss aversion task, we code the variable such that it ranges from -6 to 0, which corresponds to a λ coefficient ranging from 0.87 to 3.01, as estimated in Gächter et al. (2007).

Table 4.1: **The effect of initial investment on additional investment.**

Model	(1)	(2)	(3)	(4)
Dependent variable: Additional investment				
Initial Cost > 0	-1.380***		-1.030***	
	(0.214)		(0.168)	
Initial Cost		-0.142***		-0.104***
		(0.023)		(0.017)
Constant	-6.288***	-6.529***	1.564***	1.555***
	(0.879)	(0.888)	(0.161)	(0.161)
Controls included	Yes	Yes	No	No
N	3619	3619	3691	3691

Note: Logit panel regressions. Robust standard errors in parentheses clustered at the participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. In model (1) and (3), the initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise. In model (2) and (4), the initial costs are coded as a continuous variable *Initial Cost*. Controls are present in Model (1) and (2) and include the measures of loss aversion, waste aversion and the score in the hypothetical sunk cost questionnaire, in addition to probability of success, gender, field of study economics and the repetitions of trials in the task. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These results also indicate that, as $C_{initial}$ increases, participants are progressively less willing to make the additional investment in the continuation stage. Figure 4.2 shows that every increase in $C_{initial}$ is accompanied by a decrease in the willingness to make the additional investment. This is supported by the negative and significant coefficient of the continuous variable *Initial Cost* in Model (2) in Table 4.1 ($p < 0.001$).

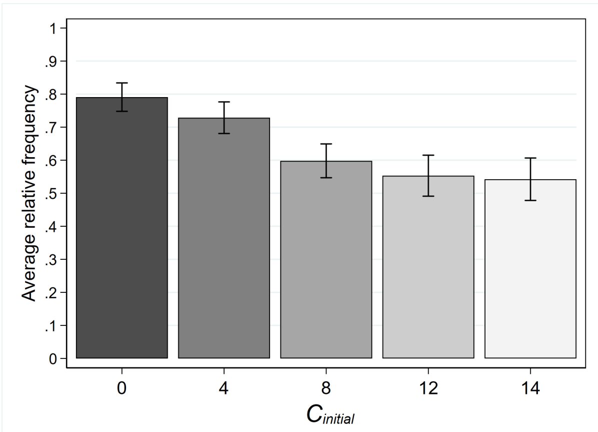


Figure 4.2: **Average relative frequency of the additional investment for each $C_{initial}$.** This figure includes both Responsibility and No Responsibility projects. The vertical bars show the 95% confidence interval, using individual averages as the unit of observation.

REVERSE SUNK COST BIAS. Participants' willingness to make additional investments decreases with $C_{initial}$.

4.2. Does being responsible for the initial investment matter?

Existing studies that investigate the sunk cost bias using monetary incentives consider only the situation where participants are responsible for the initial investment. According to the responsibility effect, they are more likely to be prone to the sunk cost bias if they are responsible for the initial investment than when someone else made this decision (Hypothesis 2).

Table 4.2: **The effect of responsibility on additional investment.**

Model: Dependent variable: Additional investment	(1)	(2)	(3)	(4)
Initial Cost > 0	-1.377***		-1.028***	
	(0.229)		(0.178)	
Initial Cost		-0.142***		-0.105***
		(0.024)		(0.018)
Responsibility	0.034	0.029	0.033	0.022
	(0.106)	(0.104)	(0.084)	(0.081)
Initial Cost > 0 × Responsibility	-0.007		-0.003	
	(0.160)		(0.118)	
Initial Cost × Responsibility		0.001		0.002
		(0.015)		(0.011)
Constant	-6.303***	-6.541***	1.548***	1.545***
	(0.878)	(0.883)	(0.169)	(0.167)
Controls included	Yes	Yes	No	No
N	3619	3619	3691	3691

Note: Logit panel regressions. Robust standard errors in parentheses clustered at the participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. In model (1) and (3), the initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise. In model (2) and (4), the initial costs are coded as a continuous variable *Initial Cost*. Controls are present in Model (1) and (2) and include the measures of loss aversion, waste aversion and the score in the hypothetical sunk cost questionnaire, in addition to probability of success, gender, field of study economics and the repetitions of trials in the task. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In *Responsibility* projects, participants made the initial investment about 90% of the time. For 99 out of the 108 participants (92%), we have at least one observation for each $C_{initial}$. In the *No Responsibility* projects the computer made the initial investment on behalf of the participants in 100% of the time. Comparing the frequency of additional investments in *Responsibility* and *No Responsibility* projects we find no difference. Specifically, when aggregating decisions over all $C_{initial} > 0$, participants are equally likely to make the additional investment in *Responsibility* and *No Responsibility* projects. In both cases, the likelihood of making the additional investment was 60.5% ($N = 107$; Wilcoxon signed-rank test, $p = 0.546$).[15] This result is corroborated in the regression analysis of Model (1) in Table 4.2 by the insignificant coefficient of the interaction between *Initial Cost > 0* and a dummy variable

[15] $N = 107$ (instead of 108) because one participant did not invest in any of the responsibility projects with $C_{initial} > 0$. We note that there is also no difference between *Responsibility* and *No Responsibility* projects if we only include those *No Responsibility* projects for which a participant decided to enter in *Responsibility* projects ($N = 107$; Wilcoxon signed rank test, $p = 0.518$). The result also holds if we exclude those who did not make an investment for all initial costs ($N = 99$; Wilcoxon signed rank test, $p = 0.5$). The same holds if we conduct the analysis separately for the two different success probabilities ($p = 0.4$: 38% in *Responsibility* and 40% in *No Responsibility*; $p = 0.6$: 83% in *Responsibility* and 81% in *No Responsibility*).

Responsibility which takes value 1 when the participant is responsible for making the initial investment ($p = 0.965$). This indicates that, when $C_{initial} >$ €0, there is no difference in the likelihood of making an additional investment between *Responsibility* and *No Responsibility* projects. Hypothesis 2a is thus rejected.

We also hypothesized that with increasing $C_{initial}$, the increase in the likelihood of making the additional investment would be stronger in *Responsibility* than in *No Responsibility* projects (Hypothesis 2b). Figure 4.3 shows how frequently the additional investment is made for each $C_{initial}$ for *Responsibility* and *No Responsibility* projects, respectively, and suggests that there is no difference between the two types of projects. This is corroborated in the regression analysis reported in Table 4.2 by the insignificant coefficient of the interaction term between *Initial Cost* and *Responsibility* ($p = 0.941$) in Model (2). As we do not find any difference in the likelihood of making the additional investment between *Responsibility* and *No Responsibility* as $C_{initial}$ increases, we thus reject Hypothesis 2b.[16]

Finally, we hypothesized that participants would be equally likely to make the additional investment in *Responsibility* and *No Responsibility* projects when $C_{initial}$ is zero (Hypothesis 2c). A Wilcoxon signed-rank test supports this hypothesis ($N = 108; p = 0.785$), which is corroborated by the insignificant coefficient of the variable *Responsibility* in Model (1) in Table 4.2 ($p = 0.747$).[17]

[16]In the linear probability model in Table 4.10 in the Appendix the coefficient of responsibility is a quite precisely estimated zero.

[17]All results regarding the additional investment discussed in Sections 4.1 and 4.2 hold (1) when considering only the participants who always make the initial investment in *Responsibility*

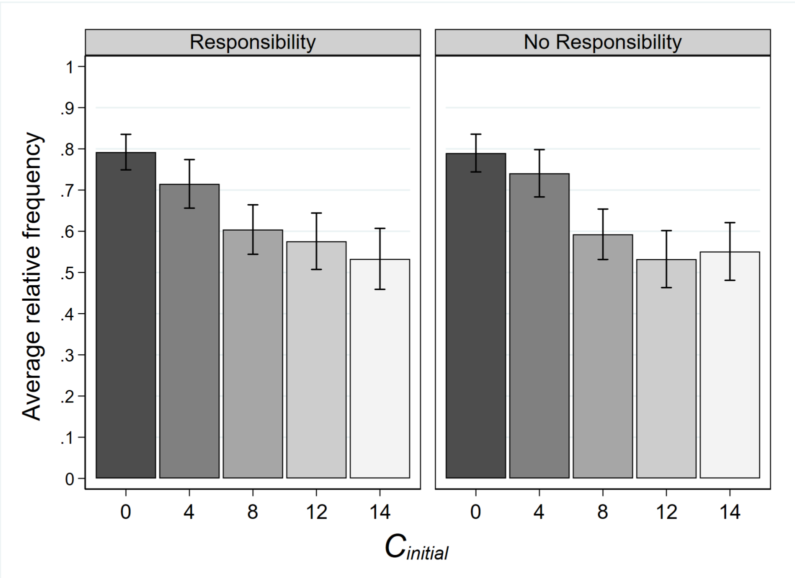


Figure 4.3: **Average relative frequency to make the additional investment for each $C_{initial}$, separately for *Responsibility* and *No Responsibility* projects.** The vertical bars show the 95% confidence interval, using individual averages as the unit of observation.

Table 4.3: **The effect of waste aversion on additional investment.**

Model: Dependent variable: Additional investment	(1)	(2)	(3)	(4)
Waste aversion	0.040^{+}	0.081^{+}	0.040^{*}	0.073^{*}
	(0.024)	(0.043)	(0.018)	(0.033)
Initial Cost > 0		0.700		0.700
		(1.473)		(1.107)
Initial Cost > 0 × Waste aversion		-0.070		-0.058
		(0.048)		(0.037)
Constant	-6.316^{***}	-7.377^{***}	-0.163	-0.629
	(0.821)	(1.287)	(0.556)	(0.970)
Controls included	Yes	Yes	No	No
N	3619	3619	3691	3691

Note: Logit panel regression. Robust standard errors in parentheses clustered at participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. The initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise. Controls are present in Model (1) and (2) and include the measures of loss aversion and the score in the hypothetical sunk cost questionnaire, in addition to probability of success, gender, field of study economics, and the repetitions of trials in the task. Results hold when controlling also for responsibility of the initial investment. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

RESPONSIBILITY EFFECT. Responsibility for the initial investment does not significantly influence the likelihood of making the additional investment.

4.3. Effects of waste aversion and of being in the loss domain

Given that we do not find evidence for a sunk cost bias, Hypothesis 3 and Hypothesis 4 on the role of waste aversion and of being in the loss domain, which were conditional on finding such an effect, are not directly tested as preregistered. However, as we find evidence for a reverse sunk cost bias, we present exploratory analyses in relation to this finding.

Model (1) in Table 4.3 tests whether higher *Waste aversion*, as measured by our questionnaire, increases the likelihood to make the additional investment. The coefficient of *Waste aversion* is positive and significant at $p = 0.091$. This suggests that participants who have a stronger self-reported waste aversion are generally more likely to make the additional investment. However, Model (2) in Table 4.3 shows that the interaction between *Waste aversion* and the dummy *Initial Cost > 0* is not significant and even has the wrong sign ($p = 0.146$). Thus, in contrast to the idea that waste aversion is an explanation for the sunk cost bias, the influence of waste aversion is not stronger for situations in which there is a "waste" in the form of sunk cost.

WASTE AVERSION. Waste aversion weakly increases the likelihood to make the additional investment. However, stronger waste aversion does not correlate positively with the willingness to make the additional investment when sunk cost is present.

To analyze the role of being in the loss domain after an initial investment, we compare projects with $C_{initial}$ of €12 and €14 with the other non-zero $C_{initial}$ of €4 and €8. In the higher initial cost projects deciding not to make the

projects and (2) when analyzing only those who not always make the initial investment in *Responsibility* projects. These robustness checks are reported in the Appendix.

Table 4.4: **The effect of being in the loss domain on additional investment.**

Model: Dependent variable: Additional investment	(1)	(2)	(3)	(4)
Loss domain	-0.779***	-1.514**	-0.567***	-1.057**
	(0.188)	(0.475)	(0.143)	(0.359)
Loss aversion	-0.315***	-0.145	-0.234**	-0.118
	(0.095)	(0.112)	(0.074)	(0.088)
Loss domain $\times$ Loss aversion		-0.329+		-0.222+
		(0.172)		(0.131)
Constant	-6.413***	-6.098***	0.314	0.569*
	(1.044)	(1.046)	(0.210)	(0.249)
Controls included	Yes	Yes	No	No
N	1612	1612	1612	1612

Note: Logit panel regression. Robust standard errors in parentheses clustered at the participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. The initial costs are coded as a dummy variable *Loss Domain* that takes value 1 when initial costs are €12 or €14, 0 otherwise. Only cases where $C_{initial}$ > €0 are included. Controls are present in Model (1) and (2) and include the measures of waste aversion and the score in the hypothetical sunk cost questionnaire, in addition to probability of success, gender, field of study economics, and the repetitions of trials in the task. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

additional investment implies to fall behind the initial endowment of €10 and therefore to be in the loss domain, whereas this is not the case for the lower initial cost projects. In Table 4.4, $C_{initial}$ is coded as a dummy variable (*Loss Domain*), which takes value 1 if $C_{initial}$ is €12 or €14. Model (1) in Table 4.4 shows that the likelihood of making the additional investment in the projects where $C_{initial}$ is €12 or €14 is lower than for the other non-zero initial cost projects ($p < 0.001$), indicating that participants rather accept to incur a certain loss relative to their initial endowment than making the additional investment. That is, instead of gambling for recovery or at least trying to get closer to the initial reference point, participants withdraw from further investments even more once they find themselves in the loss domain.

The coefficient estimate of the variable *Loss aversion* is negative and significant ($p < 0.001$), which indicates that those participants who are more loss averse are generally less willing to make the additional investment. The interaction between *Loss Domain* and *Loss aversion* in Model (2), is negative and significant at $p = 0.056$.[18] This shows that the more loss averse a participant is, the less likely he or she is to make the additional investment when $C_{initial}$ is €12 or €14. These results also hold when using the loss aversion coefficient λ as calculated in Gächter et al. (2007).

LOSS DOMAIN. Being in the loss domain decreases the likelihood of making the additional investment. Moreover, loss aversion decreases the likelihood to make the additional investment.

4.4. Sunk cost in hypothetical settings vs. incentivized settings

At the end of the experiment, our participants answered four hypothetical scenarios adapted from Arkes and Blumer (1985). This allows us to first compare the answers of our participants to those reported in the sunk cost literature and second to test whether participants' reaction to sunk cost is similar in our incentivized setting and the hypothetical scenarios. While the

[18]The distribution of the switching points and the implied loss aversion λ as calculated in Gächter et al. (2007) are shown in the Appendix

normative predictions are clear and indicate that a rational agent should ignore sunk cost, we observe that only 15 participants out of 108 (14%) answered accordingly and ignored the sunk cost in all four hypothetical scenarios. The median participant answered two out of the four scenarios consistent with the sunk cost bias, indicating a moderate tendency to be prone to the sunk cost bias in hypothetical scenarios. This tendency is stronger for the two hypothetical scenarios where participants are told to imagine they are responsible for having invested some initial cost compared to the two scenarios where they were not responsible ($N = 108$; Wilcoxon signed-rank test, $p = 0.019$).[19]

Thus, our results in the hypothetical scenarios are generally in line with the findings reported in Arkes and Blumer (1985) and we replicate them using a larger sample size. In addition, in our within-subject design in the hypothetical scenarios, we also find support for the responsibility effect as participants show a stronger tendency to show a sunk cost bias in the scenarios where they were told to imagine to be responsible for the sunk cost compared to when they were asked to imagined not to be responsible.

However, the evidence of a sunk cost bias in hypothetical scenarios does not translate to behavior in the incentivized investment task. Figure 4.4 plots the additional investment for each initial cost and each score of the hypothetical sunk cost bias (measured from 0 to 4). It shows that participants exhibit the reverse sunk cost bias in the incentivized investment task, irrespective of the strength of the sunk cost bias in the hypothetical scenarios. Even the subset of participants with a strong tendency to consider sunk cost in the hypothetical settings (3 or 4 out of 4 answers in line with a sunk cost bias) shows the reverse sunk cost pattern in the incentivized task. Thus, the tendency to take sunk costs into account measured by the hypothetical scenarios does not translate to behavior in the incentivized investment task.

HYPOTHETICAL AND REVEALED SUNK COST BIAS. We replicate the sunk cost bias using hypothetical scenarios. However, even the subjects who exhibit a strong proneness to the sunk cost bias in the hypothetical scenarios show the reverse sunk cost bias in the incentivized setting.

[19]See Appendix for the distribution of sunk cost scores in the hypothetical scenarios.

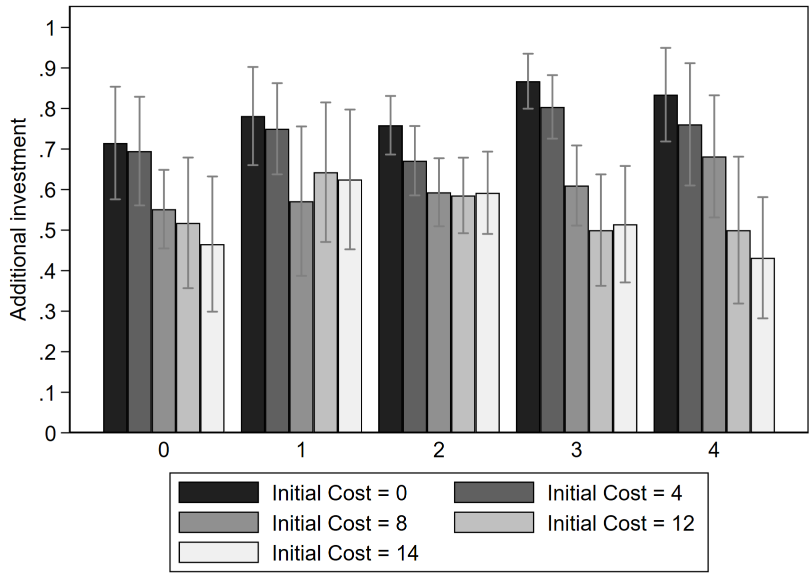


Figure 4.4: **Additional investment by hypothetical sunk cost score.** Mean decision to make the additional investment, separately for each score in the hypothetical scenario. The vertical bars show the 95% confidence interval, using individual averages as the unit of observation.

5. Discussion and conclusion

Using a novel experimental design, we study the sunk cost bias in an incentivized setting and assess potential channels underlying such an effect. We find that participants do indeed react to exogenous variations in sunk cost, but not in the direction predicted by the sunk cost bias. Instead, our main finding is that participants are less likely to make an additional investment the higher the sunk cost. Our data thus provide evidence for a reverse sunk cost bias.

The overall pattern of our behavioral findings regarding the effect of sunk cost can be rationalized in different ways. For example, a risk averse decision maker with a CRRA utility function who thinks about the payoffs from the experiment in isolation (i.e., considers only the money in the experiment in the utility function instead of her total wealth) would be increasingly less likely to make the additional investment as sunk costs increase.

A loss averse decision maker with linear utility in the loss and the gain domain could also display a pattern of behavior that resembles the one we observe. This would be the case if the reference point of the decision maker is somewhat lower than the endowment of €10. Such a reference point could reflect the amount a participant expects to (minimally) gain for the time spent on the investment task. For instance, with a reference point of €7 making the additional investment becomes a mixed gamble for initial costs of €8, €12, and €14. A sufficiently loss averse decision maker would therefore be less likely to make the additional investment compared to when initial costs are €0 and €4.

Our findings are also reminiscent of ideas suggested in two previous papers in the psychology literature. Heath (1995) argues that participants set a mental limit for their expenditures when this is easily feasible. When an expenditure reaches the mental limit, they quit investing. In a study with hypothetical incentives, Zeelenberg and Van Dijk (1997) find that behavioral sunk costs increase risk averse instead of risk seeking behavior in a subsequent monetary gamble. The authors argue that this is more likely to occur when a risk avoiding choice allows reaching an aspiration level. In our setup, an aspiration level equal to earning a positive payoff in the task could explain that participants are more willing to gamble when the "losing" outcome of the additional investment is considered satisfactory (with low $C_{initial}$), but not when the "losing" outcome fails to satisfy the aspiration level (with high $C_{initial}$). We note however, that one could also argue in favor of the opposite. If costs have been high, people could be far off their aspiration levels and therefore willing to gamble in order to reach it. Exploring which exact channel is at work would be an interesting question for future research.

Our results clearly show that findings using hypothetical scenarios do not necessarily translate into behavior with real monetary consequences. Our findings in the hypothetical scenarios replicate the ones in the seminal work by Arkes and Blumer (1985). However, even when considering only the participants who exhibit a strong hypothetical sunk cost bias, we observe a reverse sunk cost bias in the incentivized setting. We consider this as strong evidence that stated preferences elicited with questionnaires and revealed preferences using incentivized methods differ substantially, at least in the sunk cost domain.

In addition to the incentives themselves, a potential explanation for this discrepancy could be that the descriptions of the decision environments in the vignette studies leave room for interpretation. The participants might therefore interpret the incentives differently than the researchers intend and in a way that actually makes reacting to sunk cost rational. Mcafee et al. (2010) characterize a broad range of environments in which this may be the case. For example, reputational concerns could play a role. Some decision makers might believe that stopping investing and admitting that the project failed would ruin their reputation as effective decision makers (Davis et al., 1997), especially when they are responsible for the initial investment. Our results for the hypothetical scenarios are consistent with this explanation as participants are more sunk cost prone in the responsibility condition. Other misconceptions of the incentives described in the hypothetical scenarios are also plausible. In contrast to this, the incentives in our investment task are clear at every stage. Participants also make their investment decisions privately and reputational concerns are thus absent.

Studies with hypothetical decision situations have found that responsibility leads to a greater sunk cost bias (Sleesman et al., 2012; Staw, 1976). In our incentivized setting, we do not find that responsibility for the initial investment affects the propensity to make the additional investment. In that respect, a potential caveat regarding our within-subject design could be that subjects would like to appear consistent. However, this should then also hold for the hypothetical decision situations where we replicate, also within-subject, the responsibility effect found in previous hypothetical studies.

Our paper contributes to the literature on the sunk cost bias and our findings underline the need for more controlled laboratory experiments with real consequences. For instance, one avenue of future research could be to test whether a more vivid and engaging initial investment (e.g., real effort task) is a crucial component for the sunk cost bias to emerge. In addition, one could also test the role of different proximity to project completion (Conlon and Garland, 1993) and different accountability in terms of reputational loss (Fox and Staw, 1979). In any case, the search for the sunk cost bias in the laboratory needs to go on.

6. Appendix

A.1. Models with coefficients of control variables reported

In this section we present the full regression models of the main text with coefficients of control variables reported. The variable *Initial Cost > 0* is coded as a dummy variable that takes on value 1 when $C_{initial} > 0$ (that is, when sunk costs are present) and the variable *Initial Cost* is coded as a continuous variable taking on the values of $C_{initial}$. *Responsibility* is a dummy variable which takes on value 1 when the participant is responsible for making the initial investment. The variable *Loss aversion* encodes the switching point in the loss aversion task. *Waste aversion* is coded as a continuous variable that reflects the individual tendency of disliking waste or to appear wasteful. *Hypothetical sunk cost score* is a continuous variable that captures the extent to which a participant exhibits a sunk cost bias at various degrees (0–4) in hypothetical scenarios. The regression models display also the coefficient estimates for probability of success, gender (1 if male), whether the field of study is economics (1 if yes), and the repetitions of trials in the task (Period).
Table 4.5 reports the results of the logit panel regressions reported in Table 4.1 (Model (1) and (2)) and Table 4.2 (Model (3) and (4)).

Table 4.5: **The effect of initial investment (Model 1 and 2) and responsibility (Model 3 and 4) on additional investment.**

Model: Dependent variable: Additional investment	(1)	(2)	(3)	(4)
Initial Cost > 0	-1.380***		-1.377***	
	(0.214)		(0.229)	
Initial Cost		-0.142***		-0.142***
		(0.023)		(0.024)
Responsibility			0.034	0.029
			(0.106)	(0.104)
Initial Cost > 0 × Responsibility			-0.007	
			(0.160)	
Initial Cost × Responsibility				0.001
				(0.015)
Loss aversion	-0.182*	-0.185*	-0.181*	-0.185*
	(0.080)	(0.082)	(0.080)	(0.082)
Waste aversion	0.044+	0.046+	0.044+	0.045+
	(0.026)	(0.027)	(0.026)	(0.027)
Period	-0.019***	-0.018**	-0.018***	-0.018**
	(0.005)	(0.005)	(0.005)	(0.005)
Male	0.178	0.173	0.178	0.173
	(0.265)	(0.272)	(0.265)	(0.272)
Economics	0.193	0.203	0.193	0.203
	(0.266)	(0.275)	(0.266)	(0.275)
Hypothetical sunk cost score	0.128	0.131	0.128	0.131
	(0.109)	(0.112)	(0.109)	(0.112)
Probability of success	13.063***	13.427***	13.063***	13.427***
	(0.906)	(0.911)	(0.906)	(0.911)
Constant	-6.288***	-6.529***	-6.303***	-6.541***
	(0.879)	(0.888)	(0.878)	(0.883)
N	3619	3619	3619	3619

Note: Logit panel regression. Robust standard errors in parentheses clustered at the participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. In model (1) and (3), the initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise. In model (2) and (4), the initial costs are coded as a continuous variable *Initial Cost*. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.6 displays the coefficients of the variables of the logit panel regression reported in Table 4.3 in the main text that investigates the effect of the waste aversion.

Table 4.6: **The effect of waste aversion on additional investment.**

Model:	(1)	(2)
Dependent variable: Additional investment		
Waste aversion	0.040^{+}	0.081^{+}
	(0.024)	(0.043)
Initial Cost > 0		0.700
		(1.473)
Initial Cost > 0 × Waste aversion		-0.070
		(0.048)
Loss aversion	-0.166*	-0.187*
	(0.073)	(0.081)
Period	-0.017***	-0.019***
	(0.005)	(0.005)
Male	0.167	0.174
	(0.241)	(0.265)
Economics	0.172	0.192
	(0.243)	(0.267)
Hypothetical sunk cost score	0.119	0.129
	(0.100)	(0.110)
Probability of success	11.962***	13.077***
	(0.904)	(0.895)
Constant	-6.316***	-7.377***
	(0.821)	(1.287)
N	3619	3619

Note: Logit panel regression. Robust standard errors in parentheses clustered at participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. The initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.7 below displays the coefficients of the variables of the logit panel regression reported in Table 4.4 in the main text reporting the role of being in the loss domain on additional investment.

Table 4.7: **The effect of being in the loss domain on additional investment.**

Model:	(1)	(2)
Dependent variable: Additional investment		
Loss domain	-0.779***	-1.514**
	(0.188)	(0.475)
Loss aversion	-0.315***	-0.145
	(0.095)	(0.112)
Loss domain $\times$ Loss aversion		-0.329^{+}
		(0.172)
Waste aversion	0.012	0.012
	(0.028)	(0.029)
Period	-0.017**	-0.016**
	(0.006)	(0.006)
Male	0.041	0.032
	(0.303)	(0.305)
Economics	0.459	0.470
	(0.310)	(0.313)
Hypothetical sunk cost score	0.044	0.042
	(0.110)	(0.110)
Probability of success	12.761***	12.901***
	(0.909)	(0.915)
Constant	-6.413***	-6.098***
	(1.044)	(1.046)
N	1612	1612

Note: Logit panel regression. Robust standard errors in parentheses clustered at the participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. The initial costs are coded as a dummy variable *Loss Domain* that takes value 1 when initial costs are €12 or €14, 0 otherwise. Only cases where $C_{initial}$ > €0 are included. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.2. Robustness checks

In this section we present different robustness checks of the regression model investigating the role of the initial investment and the role of responsibility (Hypothesis 1 and Hypothesis 2). The main finding of a reverse sunk cost effect and the null effect of responsibility are robust and present in all the different checks reported below, such as when considering only the participants who always made the initial investment when responsible for it, when analyzing only those who not always make the initial investment when responsible for it, or when using a linear probability model instead of a logit model.

A.3. Only participants who always made the initial investment

Table 4.8 includes only those participants who always made the initial investment when responsible for it. Model (1) and (2) reports the effect of the initial investment on the likelihood to make the additional investment (cf. Table 4.1 in the main text). Model (3) and (4) report the effect of responsibility (cf. Table 4.2 in the main text).

Table 4.8: **The effect of initial investment (Model 1 and 2) and responsibility (Model 3 and 4) on additional investment.**

Model: Dependent variable: Additional investment	(1)	(2)	(3)	(4)
Initial Cost > 0	-1.305***		-1.295***	
	(0.307)		(0.317)	
Initial Cost		-0.136***		-0.131***
		(0.032)		(0.033)
Responsibility			0.088	0.130
			(0.137)	(0.132)
Initial Cost > 0 × Responsibility			-0.021	
			(0.197)	
Initial Cost × Responsibility				-0.011
				(0.018)
Loss aversion	-0.025	-0.025	-0.025	-0.025
	(0.126)	(0.130)	(0.126)	(0.130)
Waste aversion	0.071*	0.073*	0.071*	0.073*
	(0.034)	(0.035)	(0.034)	(0.035)
Period	-0.017*	-0.018*	-0.017*	-0.017*
	(0.007)	(0.008)	(0.007)	(0.008)
Male	0.385	0.395	0.385	0.395
	(0.415)	(0.428)	(0.415)	(0.428)
Economics	0.133	0.140	0.133	0.140
	(0.443)	(0.455)	(0.443)	(0.455)
Hypothetical sunk cost score	0.236	0.241	0.236	0.241
	(0.167)	(0.171)	(0.167)	(0.171)
Probability of success	13.565***	13.914***	13.567***	13.918***
	(1.458)	(1.483)	(1.458)	(1.485)
Constant	-7.095***	-7.302***	-7.142***	-7.371***
	(1.181)	(1.187)	(1.181)	(1.184)
N	2196	2196	2196	2196

Note: Logit panel regression. Robust standard errors in parentheses clustered at the participant level. Includes only participants who always made the initial investment. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. In model (1) and (3), the initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise. In model (2) and (4), the initial costs are coded as a continuous variable *Initial Cost*. This table includes only the participants who always made the initial investment when responsible for it. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.4. Only participants who did not always make the initial investment

Table 4.9 below includes only those participants who did not always made the initial investment when responsible for it. Model (1) and (2) reports the effect of the initial investment on the likelihood of making the additional investment (cf. Table 4.1 in the main text). Model (3) and (4) reports the effect of responsibility of the initial investment (cf. Table 4.2 in the main text).

Table 4.9: **The effect of initial investment (Model 1 and 2) and responsibility (Model 3 and 4) on additional investment.**

Model:	(1)	(2)	(3)	(4)
Dependent variable: Additional investment				
Initial Cost > 0	-1.474***		-1.460***	
	(0.286)		(0.328)	
Initial Cost		-0.148***		-0.154***
		(0.032)		(0.035)
Responsibility			-0.038	-0.112
			(0.160)	(0.158)
Initial Cost > 0 × Responsibility			-0.033	
			(0.253)	
Initial Cost × Responsibility				0.014
				(0.025)
Waste aversion	0.016	0.017	0.016	0.017
	(0.036)	(0.037)	(0.036)	(0.037)
Period	-0.020**	-0.018*	-0.020**	-0.018*
	(0.008)	(0.008)	(0.008)	(0.008)
Male	0.009	-0.013	0.011	-0.010
	(0.320)	(0.328)	(0.320)	(0.328)
Economics	0.083	0.091	0.082	0.090
	(0.326)	(0.340)	(0.325)	(0.340)
Hypothetical sunk cost score	0.069	0.072	0.069	0.072
	(0.138)	(0.142)	(0.138)	(0.142)
Loss aversion	-0.272**	-0.278**	-0.273**	-0.279**
	(0.092)	(0.095)	(0.093)	(0.095)
Probability of success	12.443***	12.827***	12.445***	12.821***
	(0.960)	(0.913)	(0.961)	(0.911)
Constant	-5.166***	-5.451***	-5.158***	-5.406***
	(1.093)	(1.133)	(1.081)	(1.116)
N	1423	1423	1423	1423

Note: Logit panel regression. Robust standard errors in parentheses clustered at the participant level. Includes only participants who did not always make the initial investment. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. In model (1) and (3), the initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise. In model (2) and (4), the initial costs are coded as a continuous variable *Initial Cost*. This table includes only the participants who did not always make the initial investment when responsible for it. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.5. Linear probability models

Table 4.10 presents regression results using a linear probability model instead of a logit model. Model (1) and (2) reports the effect of the initial investment on the likelihood of making the additional investment (cf. Table 4.1 in the main text). Model (3) and (4) reports the effect of responsibility of the initial investment (cf. Table 4.2 in the main text).

Table 4.10: **The effect of initial investment (Model 1 and 2) and responsibility (Model 3 and 4) on additional investment.**

Model: Dependent variable: Additional investment	(1)	(2)	(3)	(4)
Initial Cost > 0	-0.186***		-0.186***	
	(0.028)		(0.030)	
Initial Cost		-0.019***		-0.019***
		(0.003)		(0.003)
Responsibility			0.004	0.003
			(0.012)	(0.012)
Initial Cost > 0 × Responsibility			0.000	
			(0.021)	
Initial Cost × Responsibility				0.000
				(0.002)
Waste aversion	0.006+	0.006+	0.006+	0.006+
	(0.003)	(0.003)	(0.003)	(0.003)
Period	-0.002***	-0.002**	-0.002***	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.026	0.025	0.026	0.025
	(0.032)	(0.032)	(0.032)	(0.032)
Economics	0.016	0.016	0.016	0.016
	(0.033)	(0.033)	(0.033)	(0.033)
Hypothetical sunk cost score	0.016	0.016	0.016	0.016
	(0.014)	(0.014)	(0.014)	(0.014)
Loss aversion	-0.022*	-0.022*	-0.022*	-0.022*
	(0.009)	(0.009)	(0.009)	(0.009)
Probability of success	1.828***	1.830***	1.828***	1.830***
	(0.131)	(0.131)	(0.131)	(0.131)
Constant	-0.359**	-0.365**	-0.361**	-0.366**
	(0.123)	(0.123)	(0.123)	(0.123)
N	3619	3619	3619	3619

Note: Linear probability regression. Robust standard errors in parentheses clustered at the participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. In model (1) and (3), the initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise. In model (2) and (4), the initial costs are coded as a continuous variable *Initial Cost*. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.6. *Initial cost as dummy variables*

Table 4.11 shows the results of a logit panel regression testing for the effect of each initial cost on the additional investment. Each initial cost is coded as a dummy variable. The results of the comparison between levels is summarized below.

- All the estimated coefficients for each initial cost level are negative and significantly different from *Initial cost = 0* ($p < 0.001$).
- *Initial cost = 4* is significantly different from all the other Initial cost level ($p < 0.001$).
- *Initial cost = 8* is not significantly different from *Initial cost = 12* ($p = 0.122$) and from *Initial cost = 14* ($p = 0.018$).
- *Initial cost = 12* is not significantly different from *Initial cost = 14* ($p = 0.405$).

Table 4.11: **The effect of each initial investment coded as dummy on additional investment.**

Model:	(1)
Dependent variable: Additional investment	
Initial cost = 4	-0.517**
	(0.163)
Initial cost = 8	-1.448***
	(0.246)
Initial cost = 12	-1.724***
	(0.302)
Initial cost = 14	-1.862***
	(0.307)
Responsibility	0.033
	(0.070)
Waste aversion	0.045^{+}
	(0.027)
Period	-0.017**
	(0.005)
Male	0.174
	(0.273)
Economics	0.203
	(0.276)
Hypothetical sunk cost score	0.132
	(0.112)
Loss aversion	-0.186*
	(0.082)
Probability of success	13.454***
	(0.915)
Constant	-6.542***
	(0.889)
N	3619

Note: Logit panel regression. Robust standard errors in parentheses clustered at the participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. The initial costs are coded as a dummy variables for each level of initial cost. The baseline category is Initial Cost = 0. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.7. Loss aversion

Table 4.12 shows the lottery choice task used to elicit loss aversion. Lotteries 1 to 5 all have a non-negative expected value. In our experiment, most participants rejected gambles with a positive expected value (see Table 4.13). Specifically, around 67% of the participants rejected at least lottery 4, some even lottery 1 to 3. The median participant's switching lottery is 2, that is, acceptance of lotteries 1 and 2 and rejection of lotteries 3 to 6, which implies a loss-aversion coefficient $\lambda = 2$. This coefficient is slightly higher than the one reported in Gächter et al. (2007) who used the same task. Two participants have multiple switching points.

Table 4.12: Lottery choice task

Lottery	Accept	Reject
#1. If the coin turns up head, then you lose €2; if the coin turns up tails, you win €6.	O	O
#2. If the coin turns up head, then you lose €3; if the coin turns up tails, you win €6.	O	O
#3. If the coin turns up head, then you lose €4; if the coin turns up tails, you win €6.	O	O
#4. If the coin turns up head, then you lose €5; if the coin turns up tails, you win €6.	O	O
#5. If the coin turns up head, then you lose €6; if the coin turns up tails, you win €6.	O	O
#6. If the coin turns up head, then you lose €7; if the coin turns up tails, you win €6.	O	O

Note: The winning prize is €6 for each lottery and the potential loss varies from €2 to €7.

Table 4.13: Loss aversion

Acceptance and rejection behavior	Percent	Implied acceptable loss	Implied λ if $v(x) = x$
7) Reject all lotteries	4.72	< € 2	> 3
6) Accept lottery #1, reject lotteries #2 to #6	20.75	€ 2	3
5) Accept lottery #1 and #2, reject lotteries #3 to #6	41.51	€ 3	2
4) Accept lottery #1 and #3, reject lotteries #4 to #6	19.81	€ 4	1.5
3) Accept lottery #1 and #4, reject lotteries #5 to #6	8.49	€ 5	1.2
2) Accept lottery #1 and #5, reject lotteries #6	1.89	€ 6	1
1) Accept all lotteries	2.83	> € 7	≤ 0.87

A.8. Additional investment by responsibility and probability

Figure 4.5 shows the percentage of additional investment by responsibility and success probability of the project, for each initial cost. The figure shows that participants invest more often when the probability of success of a project is high (60%) than when it is low (40%) and that participants make the additional investment less often in *Initial cost* projects (i.e., Initial Cost > 0 Euros) vs. *No initial cost* projects (i.e., Initial Cost = 0 Euros). There is no difference between *Responsibility* and *No Responsibility* projects.

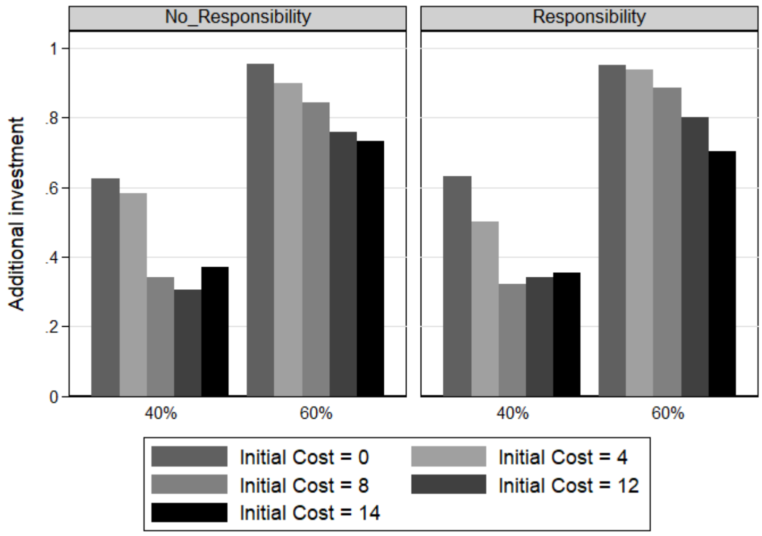


Figure 4.5: **Additional investment by responsibility and probability of success.**

A.9. Hypothetical sunk cost score

Below are the hypothetical scenarios used in the sunk cost questionnaire. We adapted vignettes 1, 3A, 8 and 9 used in Arkes and Blumer (1985). Textual changes to the original versions are emphasized in *italic*. The original text is shown between brackets.

1. *Responsibility*. Imagine that you have spent *€400* [$100] on a ticket for a weekend ski trip to *France* [Michigan]. Several weeks later you buy a *€200* [$50] ticket for a weekend ski trip to *Italy* [Wisconsin]. You think you will enjoy the *Italy* [Wisconsin] ski trip more than the *France* [Michigan] ski trip. As you are putting your just-purchased *Italy* [Wisconsin] ski trip ticket in your wallet, you notice that the *France* [Michigan] ski trip and the *Italy* [Wisconsin] ski trip are for the same weekend! It's too late to sell either ticket, and you cannot return either one. You must use one ticket and not the other. Which ski trip will you go on?

Sunk cost answers: France (45/108 participants, 41.67%; in Arkes and Blumer (1985): 64%)

2. *Responsibility*. As the president of an airline company, you have invested 10 million *Euros* [dollars] of the company's money into a research project. The purpose was to build a plane that would not be detected by conventional radar, in other words, a radar-blank plane. When the project is 90% completed, another firm begins marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and far more economical than the plane your company is building. The question is: should you invest the last 10% of the research funds to finish your radar-blank plane?

Sunk cost answer: Yes (74/108 participants, 68.52%; in Arkes and Blumer (1985): 85%)

3. *No Responsibility*. The Acme Airline Company has invested 10 million *Euros* [dollars] of the company's money into a research project. The purpose was to build a plane that would not be detected by conventional radar, in other words, a radar-blank plane. When the project is 90% completed, another firm begins marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and more economical than the plane Acme is building. The question is should Acme Airlines invest the last million Euros of its research funds to finish the radar-blank plane?

Sunk cost answer: Yes (74/108 participants, 68.52%; in Arkes and Blumer (1985): 63%)

4. *No Responsibility*. As you are listening to the radio one morning, the disk jockey calls you. He informs you that you have won a free ski trip to either *France* [Michigan] or *Italy* [Wisconsin] the last weekend of skiing season (which happens to be next weekend). You think you will prefer the trip to *Italy* [Wisconsin] rather than the trip to *France* [Michigan]. You call a travel agent and find out that the value of the *Italy* [Michigan] ski trip is €200 [$100], and the value of the *France* [Wisconsin] ski trip is €400 [$50]. You must decide which trip to take. Which trip will you go on?

Sunk cost answer: France (30/108 participants, 27.78%; in Arkes and Blumer (1985): 51%)

In our study, most people answer at least one time according to the sunk cost bias predictions. Figure 4.6 shows the percentage of people for the number of answers consistent with sunk cost bias in the hypothetical scenarios. A score of 0 means that participants never answered in line with the sunk cost bias. A score of 4 means that a participant always exhibited a the sunk cost bias. Figure 4.7 shows the percentage of people for the number of answers consistent with the sunk cost bias in the hypothetical scenarios, by responsibility. A score of 2 means that a participant always exhibited the sunk cost bias in each category (*Responsibility* and *No Responsibility*).

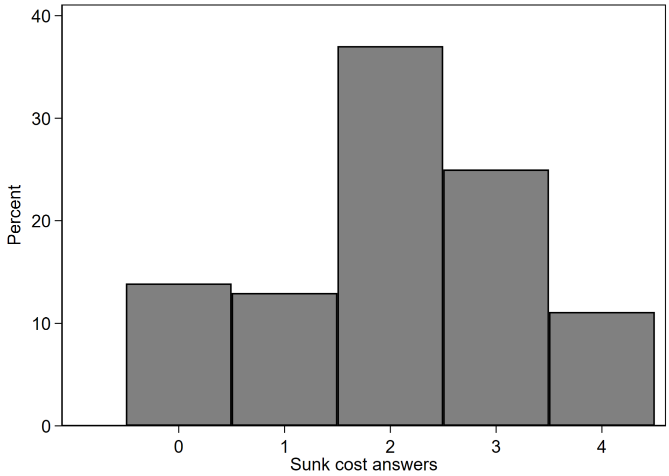


Figure 4.6: **Hypothetical scenario sunk cost**

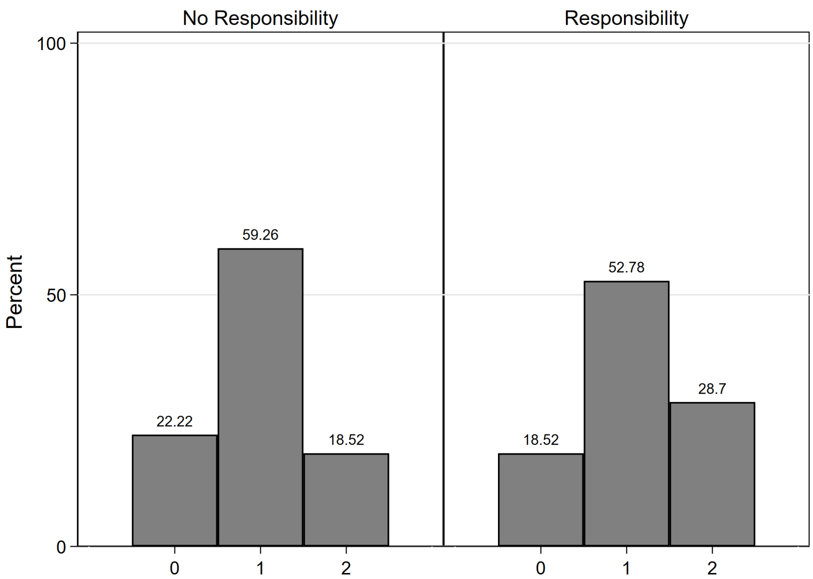


Figure 4.7: **Hypothetical scenario sunk cost by responsibility**

Table 4.14 reports logit regression results of the effect of the initial investment when considering only the participants who display a strong tendency to exhibit the sunk cost bias in the hypothetical scenarios (that is, those participants who answered 3 or 4 out of 4 answers consistent with the sunk cost bias).

Table 4.14: **The effect of initial investment on additional investment.**

	(1)	(2)	(3)	(4)
Initial Cost > 0	-1.961***		-2.232***	
	(-5.85)		(-6.16)	
Initial Cost		-0.215***		-0.234***
		(-5.47)		(-5.75)
Responsibility			-0.229	-0.121
			(-1.15)	(-0.68)
Initial Cost > 0 × Responsibility			0.563*	
			(1.98)	
Initial Cost × Responsibility				0.0410
				(1.50)
Waste aversion	-0.0152	-0.0203	-0.0159	-0.0206
	(-0.32)	(-0.39)	(-0.33)	(-0.39)
Period	-0.0167*	-0.0180*	-0.0164*	-0.0181*
	(-2.02)	(-1.99)	(-1.96)	(-2.01)
Male	0.144	0.130	0.146	0.134
	(0.38)	(0.31)	(0.38)	(0.32)
Economics	0.734+	0.819+	0.741+	0.825+
	(1.68)	(1.73)	(1.69)	(1.73)
Hypothetical sunk cost score	-0.176	-0.190	-0.175	-0.188
	(-0.40)	(-0.39)	(-0.39)	(-0.39)
Probability of success	12.64***	13.66***	12.70***	13.71***
	(10.19)	(11.22)	(10.27)	(11.22)
Loss aversion	-0.377***	-0.407***	-0.379***	-0.408***
	(-3.35)	(-3.31)	(-3.36)	(-3.32)
Constant	-3.713+	-3.964	-3.617	-3.925
	(-1.65)	(-1.63)	(-1.60)	(-1.61)
N	1287	1287	1287	1287

Note: Logit panel regressions. Robust standard errors in parentheses clustered at the participant level. Dependent variable = 1 if the additional investment in the continuation stage is made, 0 otherwise. In model (1), the initial costs are coded as a dummy variable *Initial Cost > 0* that takes value of 1 if sunk costs are present, 0 otherwise. In model (2), the initial costs are coded as a continuous variable *Initial Cost*. This table includes only the participants who answered more than 2 questions out of 4 according to sunk cost predictions in the hypothetical task. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.10. Instructions

Instructions Investment task

Welcome to this decision-making experiment.[20] You will make choices in several decision situations. All your decisions are private and you do not interact with anyone. Your earnings only depend on your own decisions and chance events. This will be explained in detail later in these instructions. If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you to answer your question in private. It is not allowed to talk, exclaim, or try to communicate with other participants during the experiment. Participants intentionally violating the rules will be asked to leave the experiment and will forfeit any earnings. This experiment consists of several parts. You will receive information on each part right before the respective part will begin. The next pages contain the instructions for the first part of the experiment. Please read the instructions carefully. You will be asked to answer comprehension questions before participating in this part of the experiment.

Your Endowment

In the first part of the experiment, you receive an initial endowment of €10. Your final earnings in this part of the experiment can be more or less than the €10 initial endowment, depending on your decisions and chance events.

Decision situations

This part consists of 36 projects. Each project, consists of up to two decision situations in which you have to decide whether you want to invest into the project or not. We call the decisions situations investment stages. For each project, there are up to two investment stages:

- An initial investment stage, where an initial investment is required for the project to start.
  - Sometimes, you can decide to make the initial investment or not.
  - Sometimes, the computer decides on your behalf to make the initial investment.

  There are costs associated with an initial investment. These costs vary across projects and can take any of these values: €0, €4, €8, €12 and €14.
- An additional investment stage. If you or the computer decides to make the initial investment, an additional investment is required to complete the project.
  - The additional investment costs are always €4.
  - You can decide to make an additional investment or not.

Initial investment stage

- An initial investment is required to start the project.
- If the initial investment is made, there will be investment costs associated with it. You will know the actual investment costs only after the initial investment is made. On average, you can expect the initial investment costs to be:
  - €0 five out of nine times;
  - €4 one out of nine times;
  - €8 one out of nine times;

[20] The text font, the size and order of images have been adapted from the original instruction version.

– €12 one out of nine times;
– €14 one out of nine times.

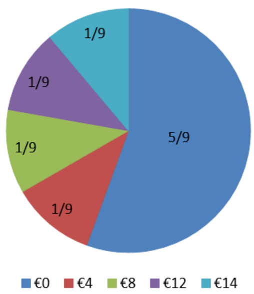


Figure 4.8: Probability of occurrence of initial investment costs, if an initial investment is made

- Sometimes, you can decide to make the initial investment or not. In this case you will see a screen like the one shown in Figure 4.9.

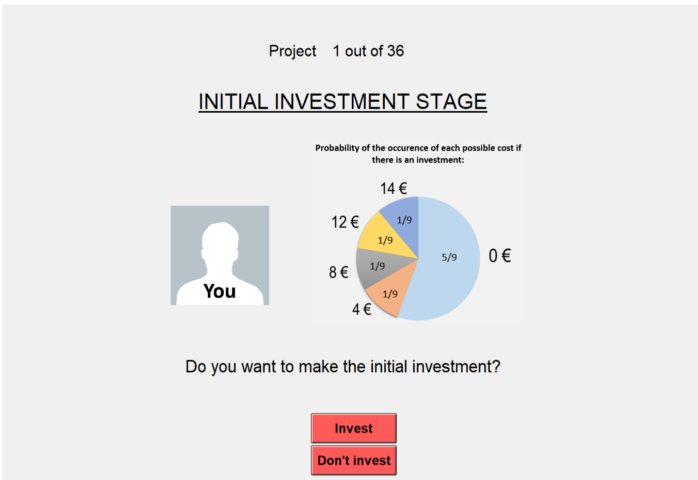


Figure 4.9: Example of a decision situation in which you can decide to make the initial investment

- After you have made the initial investment, you will be informed about the actual initial investment costs for the current project. This ends the initial investment stage.
- 4.10 shows an example where you decided to make the initial investment and you are informed that the initial investment costs are €14.
- Sometimes, the computer makes the initial investment on your behalf. In this case you will see a screen like the one shown in Figure 4.11.

Figure 4.12 shows an example where the computer made the initial investment and the costs are €14.
If you or the computer do not make the initial investment, you will move to the next project.

Figure 4.10: Example of a decision situation in which the initial investment you made is €14

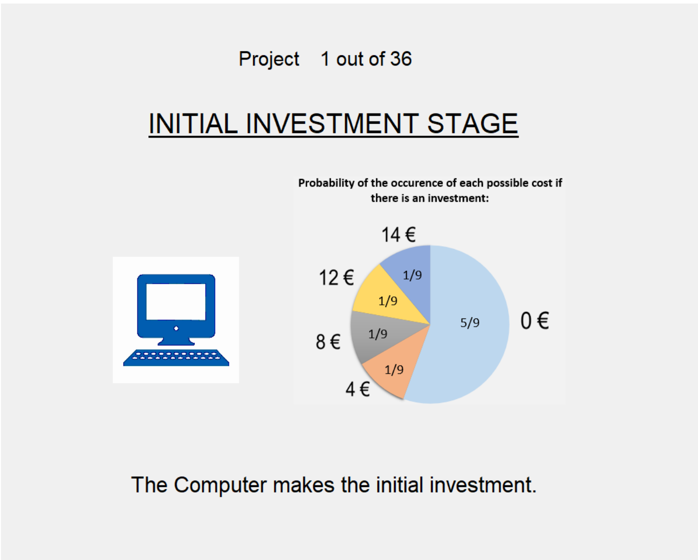


Figure 4.11: Example of a decision situation in which the computer makes the initial investment on your behalf

Figure 4.12: Example of a decision situation in which the initial investment the computer made is €14

Additional investment stage

- If the initial investment is made (by you or by the computer), you proceed to the additional investment stage.
- To complete the project, it is necessary to make an additional investment.
- The additional investment costs are always €4.
- If you make the additional investment, there is a probability that the project is successful. This probability will be either 40% or 60%.
- For half of the projects, the probability of success is 40% and for the other half it is 60%.

The example in Figure 4.13 shows a project with a probability of success of 40%.

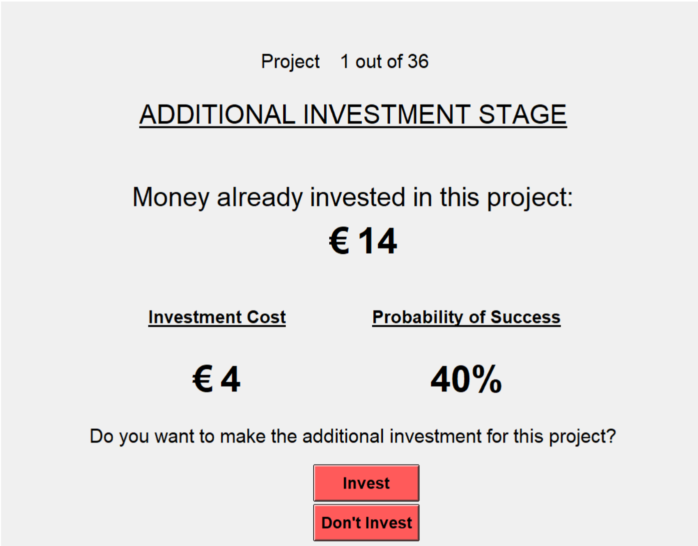


Figure 4.13: Example of additional investment. You will see the additional investment stage, only if the initial investment is made (by you or by the computer). You are reminded about the amount invested in the initial investment stage and you are presented with the additional investment costs and the probability of success of the project.

- You earn different amounts depending on whether the project is successful or unsuccessful.
- If the project is successful, you earn €16
- If the project is unsuccessful, you earn €8.
- You can also decide not to make the additional investment. In this case the project is always unsuccessful.

Your earnings after the additional investment stage:

If the additional investment is not made, the project is always unsuccessful. Your final payment in this case is:

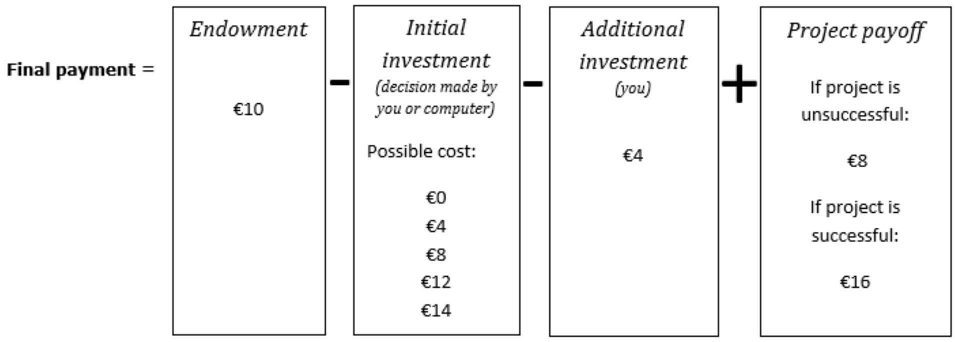
- Endowment (€10) – Initial investment + €8

If the additional investment is made, your final payment is:

- If the project is unsuccessful:
  Endowment (€10) – Initial investment – Additional investment (€4) + €8
- If the project is successful:
  Endowment (€10) – Initial investment– Additional investment (€4) + €16

Final payment calculation

Final payment = €10 Endowment - Initial investment – €4 Additional investment + Project Payoff



- If the initial investment is not made, your final payment is: *Endowment (€10)*.
- If the initial investment is made, the initial investment costs (i.e., €0, €4, €8, €12 or €14) are deducted, and you proceed to the additional investment stage.
- If the additional investment is not made, the project is automatically unsuccessful and your final payment is:
  Endowment (€10) – Initial investment + Unsuccessful Project payoff (€8).
- If the additional investment is made, you invest an additional €4 and the project can be successful according to the probability of success of 40% or 60%.
  - If the project is successful, your final payment is:
    Endowment (€10) – Initial investment – Additional investment (€4) + Successful Project payoff (€16)
  - If the project is unsuccessful, your final payment is:
    Endowment (€10) – Initial investment – Additional investment (€4) + Unsuccessful Project payoff (€8)

Your Decisions

You will face 36 decision situations. We ask you to make your investment decisions within the time limit you will see on the top right part of the screen.

Information: Until the end of the experiment you will not receive any feedback regarding the success of the projects.

Payment

At the end of the experiment, one of the 36 decision situations will be randomly selected for payout. Each decision situation is equally likely and your decisions in one decision situation do not affect any other decision situation. Thus, each decision situation could be the one relevant for your payment. Therefore, carefully consider your decisions in each decision situation. The amount you earn will be transferred to your bank account at the end of the experimental session. All bank transfers will be made on the same day of the experiment by someone who is not directly involved in this study and who does not have access to the data related to your decisions. Your anonymity is thus guaranteed.

Your Identity

For the entire duration of the experiment, your anonymity and the anonymity of all other participants is guaranteed. You will never be asked to reveal your choices to other participants during or after the experiment. Your decisions are thus private. Your name and personal information are recorded only at the end of the experiment to pay you. To keep your decisions private, please do not reveal your choices to any other participant. All the interactions are anonymous and none of the participants can communicate or influence the choices of the others.

This is the end of the instructions for Part 1. You can now answer the comprehension questions. If you have any questions please raise your hand and wait for the experimenter.

Instructions loss aversion task

Welcome to this part of the experiment. You see below a list of lotteries. Each lottery consists of a coin toss with 50% probability of success. You are asked to Accept or Reject each of these lotteries. If you Accept the lottery, you can either win or lose money. This amount will be added (if the lottery is successful) or subtracted (if the lottery fails) from the €7 you earned by answering the questionnaire in the previous part of the experiment. If you Reject the lottery, you will neither win nor lose money and you receive the €7 you earned by completing the questionnaire. At the end of the experiment, one of the lotteries will be randomly selected for payout and you will be paid according to your decision. Any earnings from this part of the experiment will be added to your previous earnings, and will be transferred to your bank account after the end of this experimental session.

[Participants then saw the lotteries as in Table 4.12].

Chapter 5

Conclusion

This dissertation presents three experimental projects that focus on the study of fairness, altruism, and sunk cost bias. I analyze these concepts from an interdisciplinary perspective, where insights from economics, psychology, and neuroscience are integrated.

In **Chapter 2**, I show that the right dorsolateral prefrontal cortex and the medial prefrontal cortex play distinct roles in punishment and reward. Despite the common belief that punishment and reward are merely two sides of the same coin, I discuss a surprising pattern of results showing distinct effects of these brain regions in reciprocal punishment and reward. The results indicate that while the right dorsolateral prefrontal cortex and the medial prefrontal cortex are broadly involved in overriding self-centered processing to facilitate the retaliation of unfairness, these same brain regions are crucial to inhibit both selfishness and reciprocity motivations when deciding how to reward kindness. The study systematizes the current knowledge on the involvement of the prefrontal cortex on reciprocal fairness. Using a within-subject design, it directly compares how punishment and reward relate to each other and provides the first evidence of the differences and similarities in the engagement of the right dorsolateral prefrontal cortex and the medial prefrontal cortex both for punishment and reward.

In **Chapter 3**, I describe a study where we experimentally examine the relationship between time delay and generosity using modified dictator games. First, I present evidence indicating that participants discount future payoff for themselves or someone else at the same rate. This is in contrast to previous findings in the literature suggesting that people discount future payoffs differently when making decisions for themselves compared to when making decisions for others. Second, no differences in aggregate giving between the four treatments are observed and, in all four treatments, men are more sensitive to the price of giving than women. Third, only when both dictator and recipient receive their payoff immediately, is giving sensitive to price changes.

In **Chapter 4**, I discuss an experimental study designed to investigate the sunk cost bias and to test several psychological mechanisms associated with it. Although the effect of sunk costs on decision-making has been widely discussed, the findings of incentivized experimental studies that investigate it are mixed. Most of the evidence in favor of the sunk cost bias comes from hypothetical scenarios of the psychological literature. It has been surprisingly hard to investigate the sunk cost bias in incentivized settings, both in the field and the lab.

In this study, we find that participants react to exogenous variation in sunk cost, but not in the expected direction. The main finding is that participants are less likely to make an additional investment the higher the sunk cost. I discuss the potential explanations of this result, and why the same participants display sunk cost bias in the hypothetical decision problems. The second main finding concerns the effect of responsibility for the initial investment. Many studies with hypothetical decision situations have found that responsibility leads to a greater sunk cost effect. In our incentivized setting, we do not observe that responsibility for the initial investment affects the propensity to make the additional investment in either direction.

Chapter 6

Impact Chapter

The research discussed in this dissertation contributes to the field of human decision sciences and includes insights relevant to psychology, experimental economics, and neuroeconomics. The studies and findings discussed can inform researchers and practitioners interested in social preferences and their biological foundations (Chapter 2), in the relationship between social and intertemporal preferences (Chapter 3), and biases in investment decisions (Chapter 4).

The study "*Love and retaliation*" presented in **Chapter 2**, provides new insights and possible future avenues in the study of reciprocity in social interactions. This study is an example of an interdisciplinary effort that has its roots in the discipline of neuroeconomics, an emerging and dynamic field that combines insights and tools from psychology, economics, and neuroscience to unravel the neural mechanisms underlying our day-to-day decisions. This interdisciplinary approach has led to insights on human decision-making that go far beyond what each separate discipline could have achieved alone. The use of innovative methods such as Transcranial Magnetic Stimulation in the study of economic relevant decisions brings an important academic contribution, by advancing the current understanding of why and when people engage in norm enforcement in the form of punishment and reward and what are the components involved in such decisions. Understanding the similarities and differences between punishment and rewards could eventually increase cooperation in society. Social punishments and rewards are important to our society to function well. However, from an academic perspective, it is still unclear whether they are just two sides of the same coin or whether they involve different psychological processes, being employed by people in different ways. In addition, for a society to work well, it is of utmost importance that social norms in place are respected and followed. Punishment and reward are tools that are used to promote norm compliance. Understanding how punishment and reward affect human behavior and how norm enforcement is encoded in the brain can bring important insights to policy-making.

The idea that motivated the study "*Intertemporal Social Preferences*" described in **Chapter 3**, is that charitable giving plays an important role in our society and understanding the donors' motivations and factors that influence giving is crucial to promote effective donations. Charitable decisions are traditionally studied in the domain of social preferences, i.e., as a decision between self and others at a certain point in time. However, in many contexts, the timing of donation and consequences to the beneficiary may not coincide or both may occur at a later point in time. For instance, when donating to a charitable organization, the costs of giving are usually immediate, but the benefits to the ultimate recipient are delayed. Also, donors could pledge to donate in the future, postponing both the cost and benefit.

These examples show that people are often confronted with situations that involve a trade-off between immediate and delayed benefits and costs (i.e., intertemporal decisions) and between one's own and others' welfare (i.e., interpersonal decisions).

To understand giving and helping behavior in such situations, it is crucial to investigate not only how people care for others and how they care for the future, but also if and how these two aspects interact. That is, the interaction between social preferences and intertemporal choice. The findings discussed in this chapter indicate that the effect of time delay on giving needs to be explored considering several factors such as different prices of giving and can inform fund-raisers and policymakers that aim at developing effective strategies to increase giving. For instance, campaigns relying on matching donations to encourage giving could thus be less efficient if they involve long time delays.

Chapter 4 discusses the study "*In search of the sunk cost bias*". This study contributes to the experimental literature that investigates decision biases. The focus is on the sunk-cost bias, which refers to the idea that makes their decision to continue an endeavor based on the previously invested resources (e.g., money, time, and effort). The sunk cost bias is considered a mistake in reasoning because sunk cost costs are costs that have already been incurred and that should thus no longer be considered relevant to future rational decision-making. Despite the intuitive appeal of the concept and a substantial body of evidence from survey studies supporting the existence of the sunk cost bias, it has been surprisingly hard to reliably demonstrate a sunk cost effect in incentivized settings, both in the field and the lab. The findings reported in this chapter are relevant to the discussion about the incentivization of participants' decisions in an experiment and shed light on the discrepancy of findings between survey and incentivized studies observed in the literature.

Public debate and promotion of Maastricht University

The work included in this dissertation has been presented and discussed in several places in Europe, America, and Asia. Among others, they include Carnegie Mellon University in Pennsylvania (2020), the European Economics Science Association Conference in France (2019) and Guatemala (2018), the Workshop on Norms and Behavioral Changes at the University of Pennsylvania (2018), the Summer school in Neuroeconomics of NYU University in Shanghai (2019), the São Paulo Summer School of Advanced Science on Social and Affective Neuroscience in Brazil (2018), the Summer School on Behavioral Economics and Psychology in Prague, and the Society for Neuroeconomics conference in Philadelphia (2018) ad Dublin (2019).

The study presented in Chapter 4 "*In search of the sunk cost bias*" was featuered in the Dutch newspaper Trouw and in the UMIO Prime engagement platform.

Software and Data analyses

The experiments presented in this thesis are programmed using z-Tree and Qualtrics. The data analyses are performed with the software STATA. The codes are available upon request to the author.

Curriculum Vitae

Marcello Negrini was born on 03 of December, 1990 in Sondrio, Italy. After graduating from the Istituto Tecnico Industriale Enea Mattei, he needed a shift in his learning and started studying psychology at Pavia University, Italy. Marcello graduated with a Bachelor of Science in Psychology in 2012 and with a Cum Laude Research Master's Degree in Psychology in 2014. He visited Vilnius University in Lithuania for six months in 2012 with the Erasmus program and stayed in Vilnius one more year for an internship in a foster home supported by the Erasmus Placement program (now called Erasmus Plus). In 2014-2015 he worked at the University of East London as a research assistant, and became a PhD candidate at Maastricht University School of Business and Economics at the end of 2015, under the supervision of Prof. Dr. Arno Riedl and Dr. Matthias Wibral. He had presented his research at various international conferences and seminars and participated in summer schools in Europe, America and Asia. In early 2020 he visited the interdisciplinary Department of Social and Decision Sciences at Carnegie Mellon University in Pittsburgh, US, before Covid-19 forced him to get back to Maastricht. Marcello will be a postdoc researcher at the Paris School of Economics starting in Fall 2021. The results of his research are presented in this thesis.

Bibliography

Afriat, S. N. (1972). Efficiency Estimation of Production Functions. *International Economic Review*, 13(3):568–598.

Agerström, J. and Björklund, F. (2009). Moral concerns are greater for temporally distant events and are moderated by value strength. *Social Cognition*, 27(2):261–282.

Albrecht, K., Volz, K. G., Sutter, M., Laibson, D. I., and Von Cramon, D. Y. (2011). What is for me is not for you: Brain correlates of intertemporal choice for self and other. *Social Cognitive and Affective Neuroscience*, 6(2):218–225.

Almenberg, J., Dreber, A., Apicella, C., and Rand, D. G. (2010). Third party reward and punishment: group size, efficiency and public goods. in Palmetti N. M. Russo J. P. (Eds.), *Psychology of punishment*. New York, NY: Nova Science.

Andreoni, J., Harbaugh, W., and Vesterlund, L. (2003). The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review*, 93(3):893–902.

Andreoni, J. and Miller, J. (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, 70(2):737–753.

Andreoni, J. and Serra-Garcia, M. (2020). Time-inconsistent charitable giving. *NBER Working Paper (22824)*.

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–56.

Andreoni, J. and Vesterlund, L. (2001). Which is the fair sex? gender differences in altruism. *The Quarterly Journal of Economics*, 116(1):293–312.

Arkes and Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin*, 125(5):591–600.

Arkes and Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision*, 35(1):124–140.

Ashraf, N., Berry, J., and Shapiro, J. M. (2010). Can higher prices stimulate product use? Evidence from a field experiment in Zambia. *American Economic Review*, 100(5):2383–2413.

Augenblick, N. (2016). The Sunk-Cost Fallacy in Penny Auctions. *The Review of Economic Studies*, 83(1):58–86.

Banerjee, R. (2016). On the interpretation of bribery in a laboratory corruption game: moral frames and social norms. *Experimental Economics*, 19(1):240–267.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., and Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, 14(11):1468–1474.

Bazerman, M. H., Giuliano, T., and Appelman, A. (1984). Escalation of commitment in individual and group decision making. *Organizational Behavior and Human Performance*, 33(2):141–152.

Bellucci, G., Chernyak, S. V., Goodyear, K., Eickhoff, S. B., and Krueger, F. (2017). Neural signatures of trust in reciprocity: A coordinate-based meta-analysis. *Human Brain Mapping*, 38(3):1233–1248.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Bogdanov, M., Ruff, C. C., and Schwabe, L. (2017). Transcranial Stimulation Over the Dorsolateral Prefrontal Cortex Increases the Impact of Past Expenses on Decision-Making. *Cerebral Cortex*, 27(2):1094–1102.

Boschini, A., Muren, A., and Persson, M. (2012). Constructing gender differences in the economics lab. *Journal of Economic Behavior & Organization*, 84(3):741–752.

Breman, A. (2011). Give more tomorrow: Two field experiments on altruism and intertemporal choice. *Journal of Public Economics*, 95(11):1349–1357.

Bridle, H., Vrieling, A., Cardillo, M., Araya, Y., and Hinojosa, L. (2013). Preparing for an interdisciplinary future: A perspective from early-career researchers. *Futures*, 53:22–32.

Brockner, J. (1992). The Escalation of Commitment to a Failing Course of Action: Toward Theoretical Progress. *Academy of Management Review*, 17(1):39.

Bronars, S. G. (1987). The Power of Nonparametric Tests of Preference Maximization. *Econometrica*, 55(3):693–698.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., and Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, 60(5):930–940.

Buckholtz, J. W. and Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, 15(5):655–661.

Buckholtz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O., and Marois, R. (2015). From blame to punishment: disrupting prefrontal cortex activity reveals norm enforcement mechanisms. *Neuron*, 87(6):1369–1380.

Cáceda, R., Prendes-Alvarez, S., Hsu, J.-J., Tripathi, S. P., Kilts, C. D., and James, G. A. (2017). The neural correlates of reciprocity are sensitive to prior experience of reciprocity. *Behavioural Brain Research*, 332:136–144.

Camerer, C. F. (2003). *Behavioral game theory. Experiments in strategic interaction.* Sage, New York.

Camerer, C. F. and Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, 97:55–95.

Cason, T. N. and Plott, C. R. (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, 122(6):1235–1270.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Chen, Y. and Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, 99(1):431–57.

Choi, B. C. and Pak, A. W. (2007). Multidisciplinarity, interdisciplinarity, and transdisciplinarity in health research, services, education and policy: 2. Promotors, barriers, and strategies of enhancement. *Clinical and Investigative Medicine*, 30(6).

Christov-Moore, L., Sugiyama, T., Grigaityte, K., and Iacoboni, M. (2017). Increasing generosity by disrupting prefrontal cortex. *Social Neuroscience*, 12(2):174–181.

Civai, C., Miniussi, C., and Rumiati, R. I. (2015). Medial prefrontal cortex reacts to unfairness if this damages the self: a tdcs study. *Social Cognitive and Affective Neuroscience*, 10(8):1054–1060.

Conlon, D. E. and Garland, H. (1993). The Role of Project Completion Information in Resource Allocation Decisions. *Academy of Management Journal*, 36(2):402–413.

Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I., and Fink, G. R. (2013). Disentangling self-and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Social Cognitive and Affective Neuroscience*, 8(4):424–431.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.

Davis, J. H., Schoorman, F. D., and Donaldson, L. (1997). Toward a stewardship theory of management. *Academy of Management Review*, 22(1):20–47.

de Kwaadsteniet, E. W., Rijkhoff, S. A., and van Dijk, E. (2013). Equality as a benchmark for third-party punishment and reward: The moderating role of uncertainty in social dilemmas. *Organizational Behavior and Human Decision Processes*, 120(2):251–259.

de Oliveira, A. C. and Jacobson, S. (2020). (Im)patience by proxy: Making intertemporal decisions for others. Working paper. *Department of Economics Working Papers 2020-02, Department of Economics, Williams College*.

Deng, Z.-D., Lisanby, S. H., and Peterchev, A. V. (2013). Electric field depth–focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coil designs. *Brain Stimulation*, 6(1):1–13.

Dreber, A., Fudenberg, D., Levine, D. K., and Rand, D. G. (2016). Self-control, social preferences and the effect of delayed payments. Working paper. *Available at SSRN 2477454*.

Easton, D. (1991). The Division, Integration, and Transfer of Knowledge. *Bulletin of the American Academy of Arts and Sciences*, 44(4):8.

Egas, M. and Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 275(1637):871–878.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

Exley, C. L., Niederle, M., and Vesterlund, L. (2020). Knowing when to ask: The cost of leaning in. *Journal of Political Economy*, 128(3):816–854.

Fehr, E. and Fischbacher, U. (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, 112(478):C1–C33.

Fehr, E. and Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4):185–190.

Fehr, E. and Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2):63–87.

Fehr, E. and Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868):137–140.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, 97(5):1858–1876.

Fox, F. V. and Staw, B. M. (1979). The Trapped Administrator: Effects of Job Insecurity and Policy Resistance Upon Commitment to a Course of Action. *Administrative Science Quarterly*, 24(3):449.

Friedman, D., Pommerenke, K., Lukose, R., Milam, G., and Huberman, B. A. (2007). Searching for the sunk cost fallacy. *Experimental Economics*, 10(1):79–104.

Gächter, S., Johnson, E., and Herrmann, A. (2007). Individual-level loss aversion in riskless and risky choices. *Institute for the Study of Labor (IZA), IZA Discussion Papers No. 2961, Available at SSRN: https://ssrn.com/abstract=1010597*.

Gächter, S., Renner, E., and Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907):1510–1510.

Gneezy, U., Imas, A., and Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science*, 60(11):2645–2658.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.

Güroğlu, B., van den Bos, W., Rombouts, S. A., and Crone, E. A. (2010). Unfair? it depends: neural correlates of fairness in social context. *Social Cognitive and Affective Neuroscience*, 5(4):414–423.

Haita-Falah, C. (2017). Sunk-cost fallacy and cognitive ability in individual decision-making. *Journal of Economic Psychology*, 58:44–59.

Haller, A. and Schwabe, L. (2014). Sunk costs in the human brain. *Neuroimage*, 97:127–133.

Heath, C. (1995). Escalation and De-escalation of Commitment in Response to Sunk Costs: The Role of Budgeting in Mental Accounting. *Organizational Behavior and Human Decision Processes*, 62(1):38–54.

Herwig, U., Satrapi, P., and Schönfeldt-Lecuona, C. (2003). Using the international 10-20 eeg system for positioning of transcranial magnetic stimulation. *Brain Topography*, 16(2):95–99.

Ho, T. H., Png, I. P., and Reza, S. (2018). Sunk cost fallacy in driving the world's costliest cars. *Management Science*, 64(4):1761–1778.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Howard, G. (2013). Discounting for personal and social payments: Patience for others, impatience for ourselves. *Journal of Environmental Economics and Management*, 66(3):583–597.

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., and Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, 45(2):201–206.

Ida, T. and Ogawa, K. (2012). Inequality aversion, social discount, and time discount rates. *International Journal of Social Economics*.

Iriberri, N. and Rey-Biel, P. (2011). The role of role uncertainty in modified dictator games. *Experimental Economics*, 14(2):160–180.

Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., and Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain*, 125(8):1808–1814.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., and Heatherton, T. F. (2002). Finding the self? an event-related fmri study. *Journal of Cognitive Neuroscience*, 14(5):785–794.

Ketel, N., Linde, J., Oosterbeek, H., and van der Klaauw, B. (2016). Tuition Fees and Sunk-cost Effects. *The Economic Journal*, 126(598):2342–2362.

Keysar, B., Converse, B. A., Wang, J., and Epley, N. (2008). Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychological Science*, 19(12):1280–1286.

Kim, H., Schnall, S., Yi, D.-J., and White, M. P. (2013). Social distance decreases responders' sensitivity to fairness in the ultimatum game. *Judgment & Decision Making*, 8(5):632–638.

Kirby, S. L. and Davis, M. A. (1998). A study of escalating commitment in principal-agent relationships: Effects of monitoring and personal responsibility. *Journal of Applied Psychology*, 83(2):206–217.

Klein, J. (1990). *Interdisciplinarity: History, theory, and practice*. Wayne State University Press.

Knoch, D. and Fehr, E. (2007). Resisting the power of temptations: the right prefrontal cortex and self-control. *Annals of the New York Academy of Sciences*, 1104(1):123–134.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 314(5800):829–832.

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., and Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proceedings of the National Academy of Sciences*, 106(49):20895–20899.

Kodaka, F., Takahashi, H., Yamada, M., Takano, H., Nakayama, K., Ito, H., and Suhara, T. (2012). Effect of cooperation level of group on punishment for non-cooperators: A functional magnetic resonance imaging study. *PloS one*, 7(7):e41338.

Kölle, F. and Wenner, L. (2018). Present-biased generosity: Time inconsistency across individual and social contexts. Working paper. *No 2018-02*.

Kovarik, J. (2009). Giving it now or later: Altruism and discounting. *Economics Letters*, 102(3):152–154.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Leibbrandt, A. and López-Pérez, R. (2012). An exploration of third and second party punishment in ten simple games. *Journal of Economic Behavior & Organization*, 84(3):753–766.

Liao, C., Wu, S., Luo, Y.-j., Guan, Q., and Cui, F. (2018). Transcranial direct current stimulation of the medial prefrontal cortex modulates the propensity to help in costly helping behavior. *Neuroscience Letters*, 674:54–59.

Loewenstein, G. and O'Donoghue, T. (2007). The heat of the moment: Modeling interactions between affect and deliberation. *Mimeo, Cornell University, 2007*, pages 1–69.

MacLeod, M. (2018). What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. *Synthese*, 195(2):697–720.

Mcafee, R. P., Mialon, H. M., and Mialon, S. H. (2010). Do sunk costs matter? *Economic Inquiry*, 48(2):323–336.

McPhee, C., Bliemel, M., and van der Bijl-Brouwer, M. (2018). Editorial: Transdisciplinary innovation (august 2018). *Technology Innovation Management Review*, 8:3–6.

Molden, D. C. and Hui, C. M. (2011). Promoting de-escalation of commitment: A regulatory-focus perspective on sunk costs. *Psychological Science*, 22(1):8–12.

Moore, D. A. and Loewenstein, G. (2004). Self-interest, automaticity, and the psychology of conflict of interest. *Social Justice Research*, 17(2):189–202.

Morse, W. C., Nielsen-Pincus, M., Force, J. E., and Wulfhorst, J. (2007). Bridges and barriers to developing and conducting interdisciplinary graduate-student team research. *Ecology and Society*, 12(2).

Nikiforakis, N. and Mitchell, H. (2014). Mixing the carrots with the sticks: Third party punishment and reward. *Experimental Economics*, 17(1):1–23.

Northoff, G. and Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences*, 8(3):102–107.

Northoff, G., Heinzel, A., De Greck, M., Bermpohl, F., Dobrowolny, H., and Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, 31(1):440–457.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, 46(8):1423–1437.

Offerman, T. and Potters, J. (2006). Does auctioning of entry licences induce collusion? an experimental study. *The Review of Economic Studies*, 73(3):769–791.

Pedersen, E. J., Kurzban, R., and McCullough, M. E. (2013). Do humans really punish altruistically? a closer look. *Proceedings of the Royal Society B: Biological Sciences*, 280(1758):20122723.

Phillips, O. R., Battalio, R. C., and Kogut, C. A. (1991). Sunk and Opportunity Costs in Valuation and Bidding. *Southern Economic Journal*, 58(1):112.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., and Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, 5(1):1–12.

Ratnadiwakara, D. and Yerramilli, V. (2017). Sunk-cost fallacy and seller behavior in the housing market. working paper. *Available at SSRN 3040712*.

Reilly, C. (2001). Transdisciplinary approach: an atypical strategy for improving outcomes in rehabilitative and long-term acute care settings. *Rehabilitation Nursing Journal*, 26(6).

Rice, M. (2013). Spanning disciplinary, sectoral and international boundaries: a sea change towards transdisciplinary global environmental change research? *Current Opinion in Environmental Sustainability*, 5(3-4):409–419.

Rogers, T. and Bazerman, M. H. (2008). Future lock-in: Future implementation increases selection of 'should'choices. *Organizational Behavior and Human Decision Processes*, 106(1):1–20.

Ronayne, D., Sgroi, D., Tuckwell, A., et al. (2020). Evaluating the sunk cost effect. *Working paper Competitive Advantage in the Global Economy (CAGE) No. 475*.

Rong, R., Grijalva, T. C., Lusk, J., and Shaw, W. D. (2019). Interpersonal discounting. *Journal of Risk and Uncertainty*, 58(1):17–42.

Ross, J. and Staw, B. M. (1993). Organizational escalation and exit: Lessons from the shoreham nuclear power plant. *Academy of Management Journal*, 36(4):701–732.

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., of TMS Consensus Group, S., et al. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, 120(12):2008–2039.

Roth, S., Robbert, T., and Straus, L. (2015). On the sunk-cost effect in economic decision-making: a meta-analytic review. *Business Research*, 8(1):99–138.

Ruff, C. C., Ugazio, G., and Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, 342(6157):482–484.

Salter, L. and Hearn, A. (1997). *Outside the lines: Issues in interdisciplinary research*. McGill-Queen's Press-MQUP.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626):1755–1758.

Schoorman, F. D. and Holahan, P. J. (1996). Psychological antecedents of escalation behavior: Effects of choice, responsibility, and decision consequences. *Journal of Applied Psychology*, 81(6):786.

Schulz, A. K. and Cheng, M. M. (2002). Persistence in capital budgeting reinvestment decisions - Personal responsibility antecedent and information asymmetry moderator: A note. *Accounting and Finance*, 42(1):73–86.

Schulz, J. F., Fischbacher, U., Thöni, C., and Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, 41:77–87.

Sefton, M., Shupp, R., and Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4):671–690.

Shapiro, J. (2010). Discounting for you, me, and we: Time preference in groups and pairs. Working paper.

Slatin, C., Galizzi, M., Mawn, B., and Melillo, K. D. (1991). Conducting interdisciplinary research to promote healthy and safe employment in health care: Promises and pitfalls. *Publich Health Reports*, 119(1):60–72.

Sleesman, D. J., Conlon, D. E., McNamara, G., and Miles, J. E. (2012). Cleaning up the big muddy: A meta-analytic review of the determinants of escalation of commitment. *Academy of Management Journal*, 55(3):541–562.

Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., and Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences*, 106(13):5035–5040.

Soman, D. and Cheema, A. (2001). The Effect of Windfall Gains on the Sunk-Cost Effect. *Marketing Letters*, 12(1):51–62.

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, 56(1):185–196.

Staw, B. M. (1976). Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance*, 16(1):27–44.

Stember, M. (1991). Advancing the social sciences through the interdisciplinary enterprise. *Social Science Journal*, 28(1):1–14.

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., and Sack, A. T. (2015). Be nice if you have to—the neurobiological roots of strategic fairness. *Social Cognitive and Affective Neuroscience*, 10(6):790–796.

Strough, J. N., Mehta, C. M., McFall, J. P., and Schuller, K. L. (2008). Are older adults less subject to the sunk-cost fallacy than younger adults?: Short report. *Psychological Science*, 19(7):650–652.

Strube, M. J. (1988). The decision to leave an abusive relationship: Empirical evidence and theoretical issues. *Psychological Bulletin*, 104(2):236–250.

Sutter, M., Lindner, P., and Platsch, D. (2009). Social norms, third-party observation and third-party reward. working paper.

Suzuki, S., Niki, K., Fujisaki, S., and Akiyama, E. (2011). Neural basis of conditional cooperation. *Social Cognitive and Affective Neuroscience*, 6(3):338–347.

Szolnoki, A. and Perc, M. (2010). Reward and cooperation in the spatial public goods game. *EPL (Europhysics Letters)*, 92(3):38003.

Tabibnia, G., Satpute, A. B., and Lieberman, M. D. (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, 19(4):339–347.

Thomsson, K. M. and Vostroknutov, A. (2017). Small-world conservatives and rigid liberals: Attitudes towards sharing in self-proclaimed left and right. *Journal of Economic Behavior & Organization*, 135:181–192.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Van Den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A., and Crone, E. A. (2009). What motivates repayment? neural correlates of reciprocity in the trust game. *Social Cognitive and Affective Neuroscience*, 4(3):294–304.

Van't Wout, M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport*, 16(16):1849–1852.

Varian, H. R. (1982). The Nonparametric Approach to Demand Analysis. *Econometrica*, 50(4):945.

Visser, M. S. and Roelofs, M. R. (2011). Heterogeneous preferences for altruism: gender and personality, social status, giving and taking. *Experimental Economics*, 14(4):490–506.

Weigel, C. (2018). Don't Stop Now: The Sunk Cost Effect in an Incentivized Lab Experiment. Working paper. *SSRN Electron. J.*

Whitfield, K. and Reid, C. (2004). Assumptions, ambiguities, and possibilities in interdisciplinary population health research. *Canadian Journal of Public Health*, 95(6):434–436.

Yi, R., Charlton, S., Porter, C., Carter, A. E., and Bickel, W. K. (2011). Future altruism: Social discounting of delayed rewards. *Behavioural Processes*, 86(1):160–3.

You, H. S. (2017). Why Teach Science with an Interdisciplinary Approach: History, Trends, and Conceptual Frameworks. *Journal of Education and Learning*, 6(4).

Zeelenberg, M. and Van Dijk, E. (1997). A reverse sunk cost effect in risky decision making: Sometimes we have too much invested to gamble. *Journal of Economic Psychology*, 18(6):677–691.

Zysset, S., Huber, O., Ferstl, E., and von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: an fmri study. *Neuroimage*, 15(4):983–991.